

Moral Sentiments in Hume and Adam Smith

Rachel Cohon, University at Albany, S.U.N.Y.

rcohon@albany.edu

Version of September 2019 for the *Oxford Handbook of Moral Psychology* ed. Doris and Vargas

Introduction: Two Challenges to Moral Sentimentalism

A sentimentalist theory of morality explains our moral evaluations of character traits and actions as manifestations of specific emotions or feelings. Following the eighteenth-century advocates of such theories, let us call the relevant emotions the *moral sentiments*. To be persuasive, such a theory must describe what the moral sentiments are like and give some account of their provenance and operation. We will examine how two eighteenth-century moral sentimentalists, David Hume and Adam Smith, attempted to do this. I will focus on interpreting their texts and noting their related yet different strategies, and I will draw attention to some difficulties they faced, ones that would have to be overcome by any contemporary version of moral sentimentalism. Indeed, their treatments of the moral sentiments have complementary strengths and weaknesses that show how difficult it can be to fulfill all the desiderata for a successful sentimentalist theory of morals.

One challenge a sentimentalist theory faces is to get the emotions right. If the theory is to be plausible, the emotions it identifies as the moral sentiments must account for, or at least be compatible with, the main features of our experience when we find or judge people or their actions to be good or evil, virtuous or vicious, right or wrong. If what happens when we feel the sentiments identified by the theory is very different from what happens when we make moral evaluations, that version of moral sentimentalism will be unconvincing. Another general task for such a theory is to describe the psychological process by which moral sentiments are

generated in us and their relation to the people and actions we evaluate. This is especially pressing for a sentimentalist who espouses naturalism and is not satisfied to say, as Francis Hutcheson does, that a moral sense is simply implanted in us by a caring deity for our ultimate benefit.¹ If finding traits and actions good or evil is a natural psychological event on a par with others, and not the exercise of a special gift of God, we should be able to explain its production in ordinary causal terms.² In identifying what the moral sentiments are like and how they are produced, one of the many important tasks for such a theory is to yield or provide some satisfactory account of the sentiments' relation to their intentional objects: how it is that these emotions are about, or directed toward, particular human agents, their character traits, or their actions. If its account of the relation between moral sentiments and their objects is over-broad, vague, or otherwise inadequate, the theory will fall short. A *cognitive* theory of moral judgments faces no such challenge (though of course it faces others); it treats moral evaluations as predications, and consequently it states very clearly what we do when we think someone is a villain, for example: we attribute a property to her. A sentiment-based theory, by contrast, must understand moral evaluation or judgment of someone in terms of emotions and their

¹ Naturalism in this context is the view that all there is in the world are natural processes and entities, that is, those that can be explained by causal laws of nature. It is common today for philosophers who see the term 'naturalism' applied to accounts of the mental to expect that the explanations offered will all be physical or material (for example neurological), but that is not intended here. Neither Hume nor Smith takes a stand on whether mental phenomena are reducible to or to be identified with physical ones. Rather, they think that all (or in Smith's case, most) mental phenomena are to be explained by natural processes, ones governed by causal laws and no supernatural intervention; but the basic items in those causal laws might be mental ones.

² To be sure, Smith is less interested in strict naturalism than Hume, and at times not only invokes final causes but also God's wisdom to explain some subtle features of our moral sentiments. See, for example, his attempted resolution of the problem of moral luck (1790 /1982, II.iii.3, especially p. 105). But the portion of his view addressed here does not appeal to any divine role in the causation of moral sentiments.

intentional objects, and this is difficult to account for in a way that maps onto the evaluations we actually make. It is hard for a theory to meet these two challenges simultaneously: to describe sentiments that match the way we in fact judge about morals, and at the same time (perhaps in the course of giving a plausible account of how they arise in us), to explain what it is for them to have the intentional objects that they have and to ensure that these are the very items we evaluate morally. Hume offers an account of what the moral sentiments are like that is in one respect faithful to our experience of moral judgment: his moral sentiment is person-evaluative rather than desire-like. And Hume gives a causal explanation of how the moral sentiments are generated from various nonmoral psychological states that is in certain ways plausible. But Hume's resulting account of how moral sentiments are of or about a person or an action proves over-broad and intuitively unsatisfying. Adam Smith gives an account of the production of a moral sentiment that is potentially more successful in pinpointing the intentional objects of those sentiments; but his account uses a less plausible model of moral evaluation, because it construes some central moral evaluations as, in significant part, desires.

I. Hume's Indirect Passions

Well before Hume writes about the moral sentiments in his *Treatise of Human Nature*, he presents a thorough, general theory of the human sentiments or passions. We begin there in order to understand the moral emotions he identifies later. In Book 2 of the *Treatise* he distinguishes between direct and indirect passions, whose difference lies in their pattern of production and in the relation between their causes and their intentional objects. Hume describes in detail four important indirect passions: pride, humility (feeling small or ashamed),

love (construed broadly, as admiration as well as liking and affection), and hatred.³ These four main indirect passions, unlike all the direct passions, have the following three distinguishing features: a) they have causes distinct from their intentional objects, b) their generation involves a shift of attention from the passion's cause to its object, and c) they take only persons, or persons understood as possessing certain qualities, as their intentional objects. For example, the indirect passion of pride is *caused* by a quality of some item that also causes an independent pleasure, such as the beauty of a certain house (which yields aesthetic enjoyment). But the sensation of pride turns my attention to, and takes as its intentional object, myself as the owner or builder of the beautiful house. In brief, the beautiful house causes my pride, but the object of my pride is myself. Humility also takes the self as object – the self as slum-dweller, for example. Love or admiration takes another person as its object, as does hatred.

This is different from the way the direct passions, which include desire, aversion, hope, and fear, operate. *Their* causes are, apparently, identical with their objects, and when direct passions arise in us they direct our attention to the same item that caused them (strongly suggested by T. 2.3.9.2 ff.), which need not be a person and quite often is not. If I fear being injured in a fall, the cause of my fear is the expectation of harm from the fall, and in feeling the fear my attention remains focused on that prospective harm. The anticipated harm is both the

³ References to Hume's work will be indicated as follows in the text: *A Treatise of Human Nature* (1739-40), ed. David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2000), is cited as T by Book, part, section and paragraph numbers. *Dissertation on the Passions* (1757), in *Four Dissertations and Essays on Suicide and the Immortality of the Soul* (South Bend, IN: St. Augustine's Press, 1992, 1995) is cited as DP by facsimile page number. *An Enquiry concerning the Principles of Morals* (1751), ed. Tom L. Beauchamp (Oxford University Press, 1988 rpt. 2004), is cited by Section and paragraph numbers.

cause of my fear and the intentional object of my fear – what I am afraid *of*. Furthermore, the direct passions, for Hume, “pursue good and avoid evil” (T 2.3.4.1); and here by ‘good’ and ‘evil’ Hume means pleasure and pain (passim., e.g., T 2.1.1.4, T 2.3.9.1). The direct passions arise from contemplating pleasure or uneasiness either experienced at present or considered in prospect, and they immediately move us to pursue the one and avoid the other when this seems feasible. Thus the direct passions are goal-directed. (To be precise, *most* direct passions arise from the thought of pleasure or pain. A few types, however, either are or arise from instincts such as hunger that are not triggered by any expectation of pleasure or pain. But these, too, are goal-directed: hunger directs my attention to food and drives me to seek it.)

The indirect passions, by contrast, while they are caused in part by pleasures and pains, do not pursue *anything*. They are not immediate motives to action, and some also do not cause any specific further motivating passions that in turn move us to act. Furthermore, the four main indirect passions (and perhaps all indirect passions) have a special kind of intentional object. They take, not pleasure or pain for the agent, nor food and drink for the agent, as their intentional objects, but rather *persons*. They are person-evaluating sentiments, as Árdal claims (1989, p. 408). We feel them toward persons (ourselves or others) for all sorts of reasons: not only because of those persons’ accomplishments or failings but because of their talent or lack of it, beauty or ugliness, even wealth or poverty. And we need not predict that the person evaluated will benefit or harm us in order feel love or hatred toward them. We love (that is, admire) the rich and powerful from a distance even when we do not stand to gain anything from them, and we love or hate heroes and villains in history and even in fiction who can neither help nor harm us. The indirect passions have hedonic tone: each of them is either

pleasant or painful. So naturally we prefer to feel pride, which is pleasant, rather than humility or shame, which is unpleasant; and similarly for love or admiration rather than hatred. But the indirect passions are not, in themselves, impulses or urges to pursue pleasure or avoid pain, or indeed to pursue or avoid anything.

For Hume, our emotional life teems with indirect passions of many sorts, many of which are not particularly ennobling, such as love of another because he is familiar, pride in one's own wealth, and even dislike of people who are ugly. They all arise naturally and are explained by Hume's associationist theory of mental processes. These indirect passions constitute a kind of *non-moral* valuing. In feeling indirect passions, we find persons (ourselves or others) good or bad in some respect, though it may be a very mundane respect.

Why would Hume expect humility and hatred, pride and love to be different in nature from desire and aversion, hope and fear, with a more complex method of generation? Why would the former passions have objects distinct from their causes? In part Hume makes this distinction in order to capture what he observes to be the differing phenomenology of these sentiments. He notices that "Pride is a certain satisfaction in ourselves, on account of some accomplishment or possession, which we enjoy" (DP 132). Usually when I feel pride I am proud of something in particular: my work, my house, my skill at dancing. But the attitude of pride is not merely enjoyment of that particular accomplishment or possession; in feeling proud I also (and especially) take pleasure in myself as a person, in the fact that I am someone with that accomplishment or possession. Similarly with humility or shame: I am of course ashamed of something particular, but that feeling of shame or low self-opinion is not merely dislike of my bad argument, my tiny cluttered apartment, or my clumsiness. It involves feeling displeased

with *myself* on their account. (One can dislike one's cluttered apartment without feeling oneself diminished by it, after all.) Hume concludes from these observations that of necessity we are aware of two different items in feeling pride and humility: the pleasing or displeasing possession or feature (the good or bad argument, lovely or ugly home) and the self. These sentiments require awareness of both. By contrast, a direct passion such as desire (for an elegant house) or aversion (to a cramped, ugly one) requires only that we be aware of the object of that desire or aversion, not that we have any special awareness of ourselves.

Another reason to regard pride and humility, love and hatred, as different from desire or aversion is that Hume conceives of all the direct passions, apart from the instincts, as aiming at pleasure or pain-avoidance for the one who feels them. Desire and aversion, hope and fear, all seek from their object pleasure or pain-avoidance for the self. They focus on what Hutcheson called natural goods and evils (1994*, pp. 70). It is the pleasure of drinking a certain kind of wine that makes the wine (non-morally) good, and I desire to drink it because (I believe) it is a source of pleasure. But Hume learned from Hutcheson that our feeling of affection for a friend is different in kind from our feeling toward a fruitful field or any other mere natural object (Hutcheson, *ibid.*, pp. 70-71). Building on this insight, Hume sees the direct passions as different in kind from person-evaluating sentiments, because the latter do not seek something for the self at all, but simply respond to features of persons and, by their response, render or constitute those persons good or bad in certain respects. The object of my pride is not some pleasure I now have or expect to get; the object of my pride is myself. In feeling it, I find myself good in a certain respect. The object of my admiration of another person is not the pleasure I may (or may not) gain from her company or services; it is the person herself, in virtue of some

characteristic of hers. In loving her (in the sense Hume has in mind, one closer to admiration), I find her to be good in a certain respect. While in feeling admiration I do feel pleasure, I assess rather than desire. Perhaps we should say that Hume's indirect passions have intentional objects but not *objectives*. They are of or about something, but that something need not be the *goal* of the one who feels them.⁴

II. Hume's Moral Sentiments

a. Some General Observations

In answer to the perennial philosophical question "do we value things because they are good, or are they good because we value them?", moral sentimentalists say the latter: actions are good, people are virtuous, and so on, because we value them, because we feel certain favorable sentiments toward them. To make this persuasive, sentimentalists must offer a psychologically plausible analysis of the attitude of valuing that does not draw on any independent standard of goodness. Some analyze valuing or approval in terms of desire, which they take to be empirically observable and understandable by science.⁵ For Hobbes (1651/1996*, Part I, ch. Iv, 25, p. 35),⁶ the account is very simple: we judge good just what we

⁴ Since we desire pleasure and we are averse to uneasiness, we can and typically do prefer the pleasure of pride to the uneasiness of humility. So while it is true that pride and humility are not themselves motives to act, for Hume ("...pride and humility are pure emotions in the soul, unattended with any [uniquely determined] desire, and not immediately exciting us to action," T 2.2.6.3), desire for the pleasure of the one and aversion to the discomfort of the other can certainly be motives to act.

⁵ In light of the on-going philosophical controversy over what, exactly, a desire is, one should not be too quick to assume that science can readily study them.

⁶ Hobbes is not a sentimentalist over all, of course, but his account of valuing fits the pattern. Spinoza, much influenced by Hobbes in this regard, puts the point very clearly: "we do not endeavor, will, seek after, or desire something because we judge it to be good, but on the contrary we judge something to be good because we endeavor, will, seek after, or desire it" (1677*, Part Three, Prop. 9, 173).

desire, so to value is to want. This may enable us to explain why valuing has an influence on action, since desires move us to act; but in other ways it accords poorly with our experience. Not all desiring is valuing and not all aversion is disvaluing; we admit that sometimes we want what we regard as bad and sometimes we do not want what we regard as good. Present-day accounts of valuing, seeking a better option, often invoke second-order desires: they may say that valuing is wanting to want something, or having a desire that I want myself to continue to have.⁷ This solves the problem of wanting something I regard as bad: I may want to play a video game all night, but I do not desire to want to do this, and so do not value doing it. But these accounts conflict with our experience of valuing in other ways. For example, sometimes we value people (morally or non-morally) without wanting anything in particular in regard to them. Various analyses of the valuing of people or traits may be suggested, such as that to value a person is to want to be like her, or to desire to desire to be like her. But this too seems to get things wrong. If I admire a great athlete of the past (if I judge her to have been *great*), there need not be anything I desire regarding her. For all my admiration of her, I don't desire to be like Florence Griffith Joyner; I don't even like running. Nor is there any desire regarding her whose persistence I in turn desire, as far as I can see. It's certainly not the case that I want to want to be like her, for example. Nor need this be true of the people we admire ethically. It is true in some cases: there are moments when I would like to have Nelson Mandela's courage and fortitude. But there are other moments when I am relieved that I don't, and am fairly content not to desire them. Yet I judge them to be great virtues – I regard them as ethically

⁷ A recent example is David Lewis (1989). But G. E. Moore criticizes such a view (without specifying its source) (1903, rpt. 1976, pp. 15-16).

excellent. These challenges facing desire-based accounts of valuing may not be insuperable. But at least at first blush, judging a person good in some respect seems quite different from wanting, and from wanting to want. A defensible moral sentimentalism needs to take these phenomena into account.

b. The Nature of Hume's Moral Sentiments

Hume identifies two moral sentiments, approbation and disapprobation. He observes that contemplating a person's virtuous character or actions in a disinterested way evokes in the observer a particular sort of pleasant emotion, and contemplating vice in an unbiased way evokes a characteristic unpleasant one. "we... must pronounce the impression arising from virtue, to be agreeable, and that proceeding from vice to be uneasy. Every moment's experience must convince us of this" (T 3.1.2.2). These feelings of approval and disapproval can be distinguished from nonmoral kinds of pleasures and displeasures in part by their unique phenomenal character, and we find that feelings of this sort are evoked only by persons and their actions and never by inanimate objects. Furthermore, only those sentiments that arise in response to our *disinterested* observations of persons, as distinct from our self-interested reactions to them, qualify as moral approval and disapproval (T 3.1.2.4). In *Treatise* Book 3 Hume does not make clear where the moral sentiments fit in his taxonomy of direct and indirect passions (if they do at all), and scholars disagree about it. I shall claim that moral

approbation and disapprobation share key features with the passions that Hume classifies as indirect, and differ from those he calls direct.⁸

The moral sentiments, as Hume describes them, have the structure of the indirect passions.⁹ Some scholars, following Árdal (1989, chapter 6), argue that the moral sentiments just *are* specialized forms of love and hatred, two of the indirect passions Hume described at length earlier in the *Treatise*. We can be agnostic about that,¹⁰ but still see that the moral sentiments are *like* Hume's indirect passions in the crucial respects. They have a cause distinct from their object. Their object is ultimately a person, though conceived as having a certain property. They turn our attention from the cause to the object. And they are not impulses of pursuit or avoidance, but rather sentiments of assessment or evaluation. In judging someone virtuous or vicious in some respect we do not feel a longing or urge to get or avoid something for ourselves, but rather we focus our attention on a person, we reflect on her quality of mind, and we feel a certain distinctive response to it. So far I assert all this dogmatically, but I will defend this interpretation when we turn to Hume's account of the causation of the moral sentiments.

⁸ For a more detailed account of this distinction and the parallels between the indirect passions and the moral sentiments, as well as discussion of some of the scholarly debates on the subject, see Cohon (2008a).

⁹ Some scholars argue that the moral sentiments, while they are of course sentiments, are not Humean passions at all, but fall outside the distinction between direct and indirect passions (see, especially, Louis Loeb 1977). This point depends in part on the claim that Hume does not treat the terms 'passion' and 'sentiment' as synonyms, and that itself is controversial. I take no stand on this issue. Whether or not they are passions, I claim, the moral sentiments have the causal and attention-directing features of the indirect passions and not those of the direct passions.

¹⁰ There are passages both strongly supporting and strongly undercutting this interpretation. See, for example, Donald Ainslie, 1999 (474-5). T 3.1.2.5 can be read either way, but (though I cannot argue for this here) I find the anti-Árdal reading more plausible.

As we have seen, Hume identifies two different kinds of non-moral passions, those that are desire-like and those that are not but instead step back and assess persons. Having given this account of the non-moral indirect passions (pride, humility, love, and hatred), Hume is now able to give an analogous account of *moral* valuing – of the feelings of approval and disapproval that constitute finding things morally good or evil. (Thus for Hume the moral sentiment is not *sui generis* in being evaluative rather than goal-directed.) On Hume’s view it is not the case that persons are good because we want them, or because we want something regarding them, or we want to be like them, or we want ourselves to want to be like them. But it is the case that they are good because we value them – because they are the objects of an indirect sentiment, one that responds in particular to persons and their character traits when we consider them in general and from the common point of view (that is, from the correct disinterested perspective for making ethical evaluations). This is a strength of Hume’s account.

c. The Causation of the Moral Sentiments

To understand the nature of Hume’s moral sentiments we need to see how they are produced in us. Hume begins with the mechanism of sympathy, by which the sentiments of someone we observe can become our own sentiments through an association of ideas and identification with the self. Sympathy is a mechanism Hume invokes in many explanations of mental occurrences apart from moral evaluation, so when he turns to explain our moral sentiments he draws on a familiar phenomenon. What Hume calls “sympathy” is a psychological process of emotion-transfer between people rather than a feeling, and for present-day readers it can be helpful to think of Humean sympathy as empathy. (The latter

word did not exist in the eighteenth century.) Let us call the person who makes the ethical evaluation – any ordinary person who makes a moral judgment – the *observer*. The observer sympathizes with the joy or suffering of some individual – call that person the *recipient*. So the joy or suffering of the recipient becomes joy or suffering in the observer through the mechanism of sympathy. The observer may also believe that the recipient's joy or suffering is caused by a quality of mind of some person, usually one that is manifested in that person's action. The quality of mind may be a trait of the recipient himself or of a third party, but for clarity let's talk about a third party. Call this third individual, whose character trait causes the recipient's joy or suffering, the *agent*. She is the person to be evaluated. Once the observer believes that the agent's character trait is the cause of the recipient's positive or negative feeling, the observer's sympathetically-acquired pleasure or pain either becomes or causes (Hume does not say which) moral approval or disapproval of the agent's character trait.

Actually, though, there is another step: at some point in this process the observer imagines herself to occupy the common point of view and sympathizes not only with the specific recipient of the harm or benefit but also with all others who are affected by the agent's mental quality, including the agent herself; and this may change the observer's over-all sympathetic feeling.¹¹ The observer's vicarious pleasure or pain, shaped by a wider sympathy,

¹¹ Reflection from the common point of view eliminates much of the bias inherent in spontaneous sympathy, which is skewed by the observer's proximity and resemblance to those with whom she sympathizes. Hume was quite concerned with the distortions that could be caused by what today is called empathy bias, and appealed to an imagined common point of view to overcome it.

either becomes or causes a feeling of approbation or disapprobation of the agent insofar as she has this character trait. (See T 3.1.2, T 3.3.1.)¹²

Here, note the cause of the moral sentiment: the recipient's weal or woe. Note its intentional object: the agent's character. The cause of the moral sentiment differs from its object. Also note that we shift our attention from the recipient to the agent – from the cause of our moral sentiment to its intentional object. We notice the suffering, for example, of the recipient, but what we disapprove of is the agent as a person who possesses a certain character trait. Also note the kind of intentional object that a moral sentiment has: a person, characterized in a certain way. We do, of course, also approve or disapprove actions, but according to Hume that feeling arises from the assumption that the action in question manifests the agent's quality of mind; and so it is a person with that quality of mind that is the primary object of our moral sentiment. I disapprove someone's unkind act, for Hume, in that I disapprove of *him* as the sort of person who would do it. Thus the moral sentiment has all three of the characteristics that distinguish indirect from direct passions. Furthermore, Hume never describes the moral sentiment as itself an impulse to act (though this may surprise some philosophers). For one thing, the moral sentiment evaluates the traits of people too distant from the observer in space and time to provide any opportunity for action. Hume does, of course, insist that our moral evaluations of ourselves "sometimes" prompt or prevent actions, at least indirectly; but he never describes moral approval or disapproval as in themselves impulses to achieve any goal. And a great many of the virtuous actions he describes are

¹² Readers who subscribe to a different interpretation of what happens when an observer consults the common point of view should substitute their own. Nothing that follows hinges on one particular interpretation of that process. The account in the text follows Cohon (2008b), ch. 5.

characteristically prompted not by moral sentiments but by other sentiments such as concern for the well-being of others.¹³ In feeling the pleasure of moral approval of another's kindness, we evaluate the person, but we need not thereby be moved to try to achieve any goal of our own. (Since moral approval is a pleasure, it is suited to play a role in motivation, just as pride is. But I would argue – though I cannot do it here – that moral approval is not itself a motive.) In this respect as well, Hume's moral approbation and disapprobation fit the pattern of the indirect, not the direct, sentiments.

If what we said above about our experience of moral evaluation is right, Hume's account of the causation of moral sentiments, their intentional objects, and their status as evaluative sentiments rather than impulses to act fits well with that experience.

d. Some Concerns about Hume's Causal Story and the Intentional Object of the Moral Sentiment

Hume's causal explanation of the genesis of the moral sentiment may seem puzzling in a certain respect. Given its origin, how does the new sentiment come to have the intentional object that it has? The observer's sympathetically-acquired pleasure or uneasiness, shaped by the common point of view, is still just a feeling of pleasure or uneasiness. It is a good or evil first caused to the recipient and others by the agent's trait that is then acquired sympathetically

¹³ See T 3.1.1, paragraphs 5 and 10. Many philosophers have the impression that Hume understands the moral sentiments to be impulses and so accepts a strong version of moral judgment internalism, but the text does not support this. So I argue in Cohon 2010. For an opposing view see Radcliffe (2018), pp. 121-137.

by the observer (T. 3.1.1 *passim.*, e.g. par.s 9-10).¹⁴ As pleasure or pain it need not have any intentional object at all. If the observer were to learn that, for example, the recipient's suffering was caused not by anyone's character trait or action but by an accident (the recipient fell and broke his collar bone, for example, resulting in distress and frustrating temporary disability), the sympathizing observer would simply feel vicarious distress and frustration, and would make no moral evaluation of anyone. That distress might have no intentional object: it might simply be vicarious anguish or sorrow. (In an instance of extreme identification with the recipient, it might even be vicarious bodily pain.) Or the distress might have as its intentional object the physical pain and disability of the recipient: the observer may be distressed *about* the recipient's pain and inability to function. Hume does not say much about the structure of the sympathetically-acquired emotion, so either is possible. Hume allows that the observer might in addition feel pity or compassion for the recipient; but this is no part of moral disapproval and plays no role in its generation. For Hume pity or compassion is a desire for the alleviation of the recipient's suffering that resembles benevolence (the instinct-based desire for the good of one's friend) but can be extended to strangers (see T 2.2.7, DP pp. 160-1). Pity/compassion is not a moral sentiment, for Hume, and it is goal-directed – it tends to move one who feels it to attempt to relieve the other's distress. But vicarious distress, perhaps abetted with compassion, do not take the *agent* (the perpetrator of the harm) as their intentional object – they can occur even if there is no agent. However, once the observer comes

¹⁴ The artificial virtues such as honesty, for example, serve the good of society, and in favorably judging someone's honesty an observer sympathizes with the whole of society in forming her approval. The natural virtues such as benevolence benefit particular individuals, and in coming to approve of one of them an observer sympathizes with those specific beneficiaries. Something parallel can be said of the artificial and natural vices.

to think the agent, a thinking and feeling person with a certain quality of mind (say, brutality and greed), caused the recipient's suffering (perhaps by pushing him down some stairs), the observer comes to feel a sentiment that is in most ways completely different from her sympathetically-induced suffering, even though it resembles it a bit in also being unpleasant. The observer shifts her attention from the distressed recipient to the agent who pushed him down the stairs, and now feels a person-evaluating attitude of disapproval of the agent. This new sentiment has an entirely new intentional object: that ill-meaning agent. The moral sentiment is about the agent: in feeling it, the observer disapproves or blames *her*. Hume does not explain how we, as observers, go from feeling vicarious pain communicated to us by one party (the recipient), and taking him (if anyone) as its object, to disapproval of someone else.

If Hume thinks of the moral sentiment of disapproval as identical to the observer's sympathetically-acquired uneasiness, then this is a serious problem for his account. On a more charitable reading, he does not actually identify these two sentiments, but instead treats the first as the cause of the second. So long as the latter is his position, Hume in fact has no problem giving a causal explanation of this transition, within the confines of his own conception of causation. For one mental state or phenomenon to cause another, or for anything to cause anything else, according to Hume, what is required is that there be a repeated pattern of association between objects or events of those two types, those of one type routinely preceding those of the other, together with a conditioned tendency of the mind, after observing the occurrence of the first type of event, to expect the second. In order for one mental state to be the cause of another, this constant conjunction and expectation is all we need. In the case of the production of the moral sentiment, both cause and effect are mental

phenomena in the observer. The first phenomenon (the cause) is a rather complex mental state: the observer's sympathetic mirroring of a recipient's distress coupled with the belief that the recipient's original distress was caused by a certain agent's character. The effect is a simpler mental state: the observer's moral disapproval of the agent. If, routinely, whenever observers experience a vicarious feeling of distress associated with a belief that a certain agent caused its original in the recipient, this feeling-belief pair is followed by a feeling of moral disapproval of that agent, and if, as a result of this frequent correlation, observers who experience a new instance of the first (complex) mental state come to expect to feel moral disapproval, then causation is established.

Now if we reflect on our experiences, we find that Hume seems to be right. When I feel vicarious distress at the suffering of another that is associated in my mind with some third party whose character I think of as the cause of that suffering, that does tend to be followed by a feeling of disapproval or blame in me directed toward the agent. The repeated pattern that Hume needs to establish this causal relation seems familiar and real.¹⁵ The moral sentiment is

¹⁵ (This note is primarily of interest to readers steeped in Hume's theory of the passions.) Hume provides a detailed account of how the four main indirect passions are caused, one that involves a double relation of ideas and impressions (the two kinds of mental items in his theory of the mind). If I am right that the moral sentiments are indirect affections, then maybe they resemble the main four indirect passions closely enough that we could reconstruct, on Hume's behalf, a causal account of the production of the moral sentiments involving an analogous double relation of impressions and ideas. If we interpret Hume as thinking that the moral sentiment just *is* a refined version of some or all of those four passions, then we will say that Hume has already provided such an account. If, instead, the moral sentiment is distinct but similar, we can construct an analogous causal story on Hume's behalf. This would tell us that the causal relation between the recipient's suffering and the agent's character joins together the ideas of these two people (recipient and agent) in the observer's mind, and this relation of ideas coupled with the resemblance between (the impressions of) vicarious suffering and moral disapprobation, insofar as both are painful, enables the first of these similar impressions (the sympathetic suffering) to call up in the observer the second impression (the moral disapprobation). The explanation would work as well for the moral sentiment as it does for pride, humility, love, and hatred. How well that works is a subject for another occasion.

very different in kind from the mental states that cause it, in that it has a different intentional object; but according to Hume's theory of causation, a cause and its effect can be as different as we please. "Any thing may produce any thing..." (T 1.3.15.1). So a sentiment that causes another need not have the same intentional object as its effect.

There are plenty of reasons to be dissatisfied with this causal account, of course. First, it is overly simple. If recipient Roland suffers bitter disappointment because he failed to win a competition for promotion, and the cause of his suffering is agent Abigail whose character trait of great ambition moved her to work much harder than her competitor to win the promotion fairly in his stead, an observer who sympathizes with Roland's disappointment need not feel any moral disapproval of Abigail. (Nor will an observer blame Arthur, supervisor of Roland and Abigail, for choosing to promote the more accomplished employee, even though Arthur is a quite direct cause of Roland's disappointment.) The account is surely too broad. Considerable refinement would be needed to make it work well, and Hume does not provide it. Furthermore, the account as stated does not exclude from moral approval and blame those we usually classify as ineligible for such evaluation. We do blame an agent who pushes a recipient down a flight of stairs out of brutality and greed, for example in order to keep him out of an athletic competition in which she has bet heavily on the opposing team. But we do not blame an agent who, in the throes of a schizophrenic "break," obeys authoritative-sounding hallucinated voices commanding her to push him. Hume does explicitly argue that moral approval and disapproval are not aroused by features of an agent that are fleeting and accidental ("casual," in one eighteenth-century sense of that term), such as a momentary loss of balance from tripping on a

rug that causes an agent to fall against the recipient; to exclude such accidents, Hume claims that the moral sentiments respond only to an agent's settled traits (T. 2.3.2.7). But serious mental illnesses and defects can be quite settled features of an agent, and yet they tend not to trigger moral disapproval. So there is, again, at least a problem of detail to be solved, and the devil may lurk in the details.¹⁶

But here let us attend to a different problem about intentional objects that is left us by Hume's causal account and that I suspect cuts deeper.

In his taxonomy of the contents of the mind, Hume classifies our sentiments, including all the direct and indirect passions, as simple impressions: impressions (original mental experiences) with no parts that can be distinguished in analysis. The various emotions and feelings differ from one another not by what parts they possess but by their phenomenological qualities and (more prominently) by their causes and effects. This leaves him in some general difficulty about how to account for the fact that our emotions are about something or directed toward something. Many different fears feel similar to one another, so what distinguishes the fear of falling down the stairs from the fear of a venomous snake or of war? Clearly, it must be the intentional object of the passion: they are fears *of* different things. Present-day

¹⁶ The claim that the moral sentiment only responds to settled traits goes too far in another way, ruling out not only accidental behavior but intentional actions done out of character. But it is worth considering how we evaluate an agent when she acts out of character in a way that elicits a strong moral response. If we learn that a benign, good-natured celebrity has murdered his wife, of course we condemn the murderous act; but we also tend to infer that he was not so benign and good-natured as we had supposed. Perhaps Hume claims that the moral sentiments only respond to settled character traits in part because we, as observers, find it very difficult to retain our prior assessment of someone's character trait in the face of an action utterly unlike those to which such a trait would give rise. (Thanks to Manuel Vargas for pressing me to address this issue.)

philosophers tend to assume almost automatically that emotions have a proposition-like structure into which is incorporated a representation of the emotion's intentional object, so that fear of falling and fear of a snake are different emotions in part because they contain within them representations of those different objects, the anticipated fall on the one hand and the snake on the other. But Hume rejects this conception and even denies that passions are or contain any representations at all (T 2.3.3 and 3.1.1).¹⁷ He nonetheless has no doubt that passions have intentional objects; but he explains them entirely in terms of causation and the direction of our attention.¹⁸ For Hume the object of a passion is that to which it directs our attention, that idea that the passion causes to arise in the mind (for example, T 2.1.2.4). If I am angry at Andrew, on Hume's account my anger's being directed at Andrew comes to the fact that (once the right sort of cause has produced anger in me) the feeling of anger causes me to think of Andrew. If I am proud of my beautiful house, that emotion of pride is directed at myself, and what that directedness comes to is that (once the right set of associated impressions and ideas has caused in me a specific sort of pleasant impression called pride) I feel

¹⁷ There is considerable controversy among interpreters about how seriously Hume means his claim that passions are not representations, or why he says such a thing. On my view, it is a consequence of his very parsimonious reduction of the contents of the mind to impressions and ideas only. Clearly passions are felt experiences, and so must be impressions. We form idea-copies of them in memory and can use those ideas (the idea of fear, the idea of pride) in thinking about our passions. But since the passions themselves are impressions, Hume lacks the means to understand them as having ideas (of their objects) as parts of the passions themselves. Impressions in general can have parts, but on Hume's account their parts can only be simpler impressions. Only ideas, however, can serve a representative function, for Hume, since (at least on his official position) to represent something is to be a copy of it, and ideas are copies of impressions, while impressions are not copies of anything, or at least anything mental. Since passions cannot have ideas as parts, therefore they cannot be or contain representations of objects.

¹⁸ Here I follow the interpretation in Cohon (2008a), influenced by Árdall (1998). For a thorough discussion of opposing interpretations, and an interpretation that harmonizes with this one but differs in some ways, see Radcliffe (2018) ch. 4.

pride and that feeling causes me to think of myself. But this account of intentional objects, while ingenious at avoiding placing any representation of the intentional object within the sentiment itself, fails to satisfy. For, to step outside the boundaries of Hume's theory for a moment, it does not sound at all adequate to say that to be angry at Andrew is the same thing as to feel anger and think of Andrew, or even that it is the same as to feel anger that causes me to think of Andrew. (If Andrew lied to me while wearing a knit cap, and when I learn he was lying I become angry at him, my feeling of anger might cause me to think of Andrew's knit cap as well, but I am not angry at the cap.) This causal relation seems too crude to be the right relation between an emotion and its intentional object.

Given Hume's general view of the intentional objects of sentiments, his account of what makes a moral sentiment one of approval or disapproval *of* someone in particular must be understood as follows. Through sympathy I acquire your suffering, and I form a cause-and-effect association of your suffering with the agent who harmed you. These associated impressions and ideas cause me to feel a new unpleasant sentiment called moral disapproval; and moral disapproval in turn causes me to think of the agent who harmed you. So what makes my disapproval a disapproval *of Andrew* is that my feeling was caused in the right way for such a feeling, and when I feel it, it makes me think of Andrew. But surely, feeling an unpleasant emotion that causes me to think of Andrew is not the same as disapproving or blaming Andrew. At best it is necessary that the feeling make me think of Andrew; but that is surely not sufficient for intentionality.

Moral blame or disapproval targets its object, is directed toward it, or takes it *as* an object, in some more specific way that Hume's causation-attention story does not fully

articulate. Otherwise, it seems, if Andrew was wearing his knit cap when he harmed someone and the moral disapproval that this generates in me causes an image of Andrew's knit cap to arise in my mind, I also feel moral disapproval of the cap, which can't be right. Or if the object of moral disapproval is artificially restricted to thinking human beings, my feeling of disapproval might cause me to think of Andrew's sweet innocent parents, although I do not morally blame them. (The parents even stand in a causal relation to Andrew, and so to his deceitfulness and misdeeds.) Whatever it is that is missing from Hume's account of the intentionality of all the other sentiments is also missing from his account of the intentionality of the moral sentiments. But it is particularly noticeable with regard to the moral sentiments, since feeling them is supposed to be identical with making ethical evaluations of persons' characters and actions. The purely causal story Hume gives us of what determines the intentional objects of our moral sentiments results in an unsatisfactory version of moral sentimentalism.

III. Adam Smith's Moral Sentiments, Particularly the Sense of Merit and Demerit

a. Their Causation

Adam Smith, in his *Theory of the Moral Sentiments*, offers a sentimentalist account of moral judgment that is influenced by Hume's but diverges from it significantly. First let us examine Smith's moral sentiments, how they are caused, and what is implicitly involved in their having the intentional objects that they do. (We consider later whether the moral sentiments as Smith describes them accurately capture what moral evaluation is really like.)¹⁹

¹⁹ Further references to Adam Smith, *The Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. Macfie (Indianapolis: Liberty Press, 1790 /1982) are given parenthetically in the text as TMS by Part, section, chapter (where relevant), and page number.

Smith has a more complex theory of the moral sentiments than Hume has. There are two distinct types of moral approbation and disapprobation, for Smith: one that evaluates agents' emotions and actions for their *propriety*, and one that evaluates them for their *merit or demerit*. These are different types of feelings – the sense of propriety and the sense of merit – and they are caused in different ways. Both moral sentiments evaluate “the ... affection of the heart, from which any action proceeds, and upon which its whole virtue or vice depends” – that is, both evaluate the agent's motivating sentiment or desire in order to judge her action. But that affection of the heart “may be considered under two different aspects” (TMS II.1, p. 67), propriety and merit. When we approve or disapprove any feeling, and any action it may prompt, for its propriety or impropriety, we approve or disapprove “the proportion or disproportion, which the affection seems to bear to the cause or object which excites it.” This is a kind of suitability relation between the situation that caused the agent's sentiment and that sentiment itself. For example, if I feel terrified of nuclear attack, since that fear is proportional to the prospect of nuclear war that excites it, an observer will feel the sense of propriety toward my fear (deem it appropriate or graceful); while if I am terrified of a mouse, since that fear is disproportional to the cause that excites it (the mouse), an observer will feel the sense of impropriety toward my fear. When, instead, we experience the sense of merit or demerit toward someone's sentiment, we experience “a distinct species of approbation and disapprobation” in response to “the beneficial or hurtful *effects* which the affection proposes or tends to produce” (ibid., my emphasis). So if I am power-hungry, which tends to have a harmful effect on others, an observer will consequently feel a sense of demerit toward that motivating

sentiment of mine; and if I am kind-hearted, which tends to produce benefit, an observer will feel the sense of merit toward that emotion of mine, approving it as meritorious.

For Smith, the moral sentiments start from sympathy, as they do for Hume. (Smith and Hume understand sympathy somewhat differently from one another, and there is controversy about their differences, though the differences do not matter here.²⁰) The moral sentiment of propriety is caused as follows. We observers imagine ourselves in the other person's place as completely as we can, imagining that we are in his circumstances and have his characteristics, and as a result we experience the emotions that we imagine we would feel ourselves if we fully occupied the other's place. (I do not mean that we do this intentionally or knowingly; but this is what occurs in the course of moral evaluation.) We automatically compare that sympathetically-acquired feeling with the recipient's actual feeling (as revealed to us by his behavior, presumably). If there is a fairly close correspondence between his feeling and our own imaginatively- and so sympathetically-acquired emotion, we experience a (separate) pleasure in the matching (in the "concordance"), or more accurately in the sharing, of the two sentiments – in the fact that we and the recipient feel just the same way. This happens even if the two sentiments that match are not themselves pleasant but instead painful; the match itself gives us pleasure (footnote to TMS I.iii.1.9 added in ed. 2, p. 46). That pleasure in sharing the other's emotion is our approval of his sentiment as *appropriate* to its triggering cause in his circumstances. This process is slightly altered once the observer has cultivated an impartial spectator within him: at that point in our development as observers, we spontaneously take up the perspective of an impartial spectator, an imaginary neutral party with no stake in the

²⁰ See, for example, Samuel Fleischacker (2012).

matter but with a vivid imagination of the emotions of others, and come to feel, ourselves, what the spectator would feel in the recipient's circumstances. (Since introduction of the impartial spectator changes little that is relevant here, I omit mention of it in what follows.) So if the recipient has broken his collarbone and consequently suffers distress, I as an unbiased observer, knowing that a broken collarbone is quite painful and temporarily disabling for anyone, come to feel the degree of distress that any neutral (but emotionally responsive) party would feel in just this situation. I automatically compare this feeling with the recipient's actual feeling and see whether I feel just as he does. If I can "bring home to myself" the recipient's suffering, then I feel a pleasure (quite separate from that distress) in the "concordance" of my sentiment with his, and in feeling this I approve the recipient's sentiment as proper and graceful. But if I find that his distress is far greater or less than what I can feel by suitably-structured sympathy, I disapprove his distress as excessive or deficient (ibid.).²¹

This is how we come to feel moral approval of the propriety both of recipients' emotional *reactions* to what befalls them and of the passions that *move* agents to act in their circumstances. If I can come to share the sentiment that moved an agent to do what she did in her circumstances, I will approve her motive, and consequently her action, as appropriate. If I

²¹ This is not an ideal example for exhibiting deficiency, because of Smith's views about our approval of self-command: feeling less than an ordinary person would is sometimes admirable, for Smith. So to illustrate disapprobation where the observer feels more vicarious distress than does the actual recipient, consider another example: someone who feels no grief at the death of a perfectly adequate father or mother. A neutral but sympathetic observer, learning of the recipient's bereavement, would feel some grief, and consequently would not be able to sympathize with the recipient's blasé response once he became aware of it; so the observer's sympathetically-acquired feeling would not accord with the recipient's actual lack of feeling, and consequently the observer would disapprove that lack of feeling as improper. Note that for Smith there is ordinarily no effort or intentional action involved in this sequence of events and we are not normally conscious of it. This is Smith's reconstruction of the psychological, causal processes that result in our feelings of propriety or impropriety.

cannot come to share her sentiment even after careful use of my imagination, I will disapprove it and its resulting action as inappropriate.

But besides the sense of propriety, we also have the sense of merit and demerit, which responds to motivating emotions and their resulting actions on the different ground I mentioned: their expected beneficial and harmful effects or tendencies. The sense of merit and demerit is also a kind of moral approbation and disapprobation, recall, but a different kind.²² According to Smith, for an observer to feel the sentiment of merit toward an agent she must sympathize with two different emotions (often in two different people), and her responses are built upon the sense of propriety. Smith calls the sense of merit a “compounded sentiment... made up of two distinct emotions; a direct sympathy with the sentiments of the agent, and an indirect sympathy with the gratitude of those who receive the benefit of his actions” (TMS II.i.5.2, p. 74). Suppose a traveler is stranded in a foreign city and an acquaintance kindly gives him a bed for the night. The traveler consequently feels pleasant relief. When I, in my role as a neutral spectator, evaluate the agent (the kind acquaintance in this case), my emotional response to her depends not only on the beneficial effects of her action but also on the motivating sentiment from which she acted (the “affections of [her] heart”). Since she acted from kindness, which is what an emotionally responsive and unbiased agent would feel in her

²² Certain passages in TMS may suggest that the only moral approval and disapproval Smith acknowledges are the feelings of propriety and impropriety. The first two or three paragraphs of TMS I.i.3 and a remark in the famous footnote to TMS I.iii.1.9 suggest this idea. Such an interpretation would not regard the sense of merit as a whole as a (separate kind of) moral approval, but rather as a distinct sentiment that incorporates within it a feeling of moral approval (i.e., a feeling of propriety). But reading even these passages in context, plus considering many others (such as TMS II.i.1.intro.1, p. 67, where he calls the feelings of merit and demerit “a distinct species of approbation and disapprobation”), shows that Smith in fact regards the sense of merit and demerit as moral sentiments – as feelings of moral approval and disapproval in their own right.

circumstances, I can fully sympathize with her motive, and so I approve her motive as *appropriate*. This is the direct sympathy. This in turn enables me to sympathize “indirectly” with something the *recipient* feels, his *gratitude* to his benefactor, which I (from an indifferent bystander’s perspective) likewise find appropriate (TMS II.ii.1-2. p. 69; TMS II.ii.4, p. 70; TMS lii.5.1, p. 74). Had I found the benefactor’s motives to lack propriety, I would “have little sympathy with the gratitude of the person who receives the benefit” (TMS II.iii.1, p. 71). This might happen if, for example, the benefactor did not act from kindness but from a desire to exploit the traveler later, or even from some silly impulse (perhaps to host someone whose hair matches the bedspread). When I do find the agent’s motive to be proper, I am able to sympathize indirectly with the recipient’s gratitude; and my sympathetically-acquired feeling of gratitude to the agent for her kindness just *is* my sense of the agent’s merit. Had the agent’s motive proved to be improper, I would not have shared in the recipient’s gratitude (even if the recipient, misguidedly, did feel grateful), and so I would not have felt the agent to be meritorious.

Much the same happens with the sense of demerit, *mutatis mutandis*. Return to the example of the recipient in distress over a broken collar bone. Suppose I learn that the cause of his fall was a deliberate push by someone who wished to hurt him for her own advantage (again, she had bet against him in an upcoming sports competition). If I consider the situation as a neutral spectator, I cannot share the motive that the agent felt; I cannot acquire by sympathy her selfish brutality.²³ Either I simply fail to sympathize with or I feel actively hostile

²³ Not only does my sympathy fail in this situation, but Smith says that as a neutral observer I may feel “antipathy” to the agent’s motive, particularly if the agent’s action-prompting sentiments are not only

to the agent's motive in this situation; so I disapprove it as improper. In part because of this, I can indirectly (but fully) sympathize with the recipient's *resentment* of the agent, which I therefore find appropriate. I actually *come to resent the agent myself* for what she did to the recipient.²⁴ This feeling of resentment is my disapproval of the agent's action and character – my sense of her demerit, a moral sentiment.

b. Contrast between Hume's and Smith's Causal Stories and their accounts of the directedness of moral sentiments

Hume's theory of the moral sentiments has no analog of Smith's sense of propriety.²⁵

But Hume's single type of moral sentiment, that person-evaluating feeling of approval or

ones I cannot enter into myself but ones I find detestable or abhorrent (TMS II.i.5.4, p. 75; and II.i.5.6, p. 76).

²⁴ John McHugh (in private correspondence) rightly points out that the progression of steps must be a bit more complex than this if we are to explain what causes antipathy to the agent's motive. The abhorrence of the agent's motive is not a separate phenomenon, but (presumably) a consequence of the fact that the observer not only fails to sympathize with the agent's motive but also does successfully sympathize with the recipient's resentment.

²⁵ (Here I try to ward off a possible misunderstanding.) For Hume, the single type of moral approbation turns out to be felt toward character traits that are either useful or immediately agreeable to their possessor or to others (see T 3.3.1.30 and EPM 9.1-3). This *might* lead some readers to think that Hume's category of the agreeable corresponds in part to Smith's category of propriety. But this would be a serious mistake. In saying that we approve traits that are useful, Hume summarizes the process described above whereby our sympathy with the beneficiaries of those traits causes our moral approval of them. Some traits, though, such as wit and military glory or heroism, may not be especially advantageous in the long run either to their possessor or to those with whom she interacts, but they generate some immediate pleasure either for other people or for the person herself. We do approve these as virtues as well, Hume claims, and our moral approval of them is caused by their immediate agreeableness. So for example, heroism or military glory in fact creates extensive suffering ("the devastation of provinces, the sack of cities"), but the contemplation of it is so immediately agreeable to those who encounter the hero ("so elevates the mind," T. 3.3.2.15) that they cannot help but approve it as a virtue, and neither can more distant observers who sympathize with that reaction. Clearly this is a different concept from Smith's notion of propriety-approval, which results from our recognition of a concordance of the emotion of an agent or patient with the feeling of the observer once he imaginatively projects himself into her situation. In finding another's sentiment or trait agreeable in the Humean sense we do not seek or find any such concordance. Furthermore, the feeling of pleasure Hume

disapproval, is somewhat like Smith's sense of merit and demerit, because it responds to the good or harm that a character trait provides to those it affects. Hume's account of the generation of moral approval and disapproval, as we saw, goes as far as his theory of causation requires to show that a mere feeling of pleasure or uneasiness generated by sympathy with one person is the cause of a very different kind of feeling, one directed to a different person and taking her as its intentional object. And that might be fine, if it could be suitably fine-tuned. But Hume's view of the passions leaves us with a poor understanding of what it is for the moral sentiment to take persons and their actions as its intentional objects. The nonrepresentational, causal understanding of intentionality lacks some needed distinctions. So we are left wondering: what is the nature of the person-directedness of moral sentiments? What more is there to it besides feeling a pleasant or unpleasant sentiment that causes us to think of someone?

Smith's account of the origin of the sense of merit and demerit promises to do a better job of capturing the way in which a moral sentiment is about a person or directed to her. Smith does not offer an explicit account of the intentionality of this or any emotions; but it is safe to assume that he does not take on Hume's thin causal account of intentionality. What he does say suggests a different picture.

An observer's feeling of the sense of merit on a particular occasion, for Smith, has quite a few parts: sympathetic pleasure acquired from observing the recipient, a sense of propriety toward the recipient's pleasure (resulting from the concordance of the observer's sentiment

describes in contemplating the hero or the witty person is not itself a special kind of moral approbation, but rather a precursor and cause of moral approbation (on the most charitable interpretation of Hume), whereas for Smith the sense of propriety is itself a kind or species of moral approbation.

with the recipient's sentiment), sympathy with the *agent's* motive and the feeling of the propriety of that motive, and finally, the vicarious sensation of the recipient's gratitude toward the agent. We can ask whether each of these has an intentional object, if so whether it is the right sort of object for moral approval, and how Smith might account for the feeling's being directed to that object. The vicarious feeling of pleasure might have no intentional object, or might be directed to some outcome the recipient appreciates (for example, the recipient may feel glad about having a place to spend the night after all); so this sentiment will not take as its object the agent's trait or action, as moral approval must. The first feeling of propriety approves the recipient's pleasure (which accords with the observer's), so this too fails to take the *agent's* sentiment or action as its object. The observer's sympathetic mirroring of the agent's motive does not have the right object either, since that is some sort of motivating feeling about a situation the agent wishes to change, such as the agent's desire to assist the recipient. The second sentiment of propriety (propriety-type approval of the agent's motive) does seem to have the agent's motive as its intentional object, though Smith offers no explanation of what this relation consists in. (How is it that the pleasure the observer takes in the concordance of his sentiment with that of the agent can be about – can be approval of – the agent's motivating sentiment? What the observer takes pleasure in is the matching of the two sentiments, so it is not clear just how this pleasure constitutes approval of the agent's sentiment, rather than approval of the concordance. But Smith appears to treat it as approval of the agent's sentiment, which at least would be the right intentional object for moral approval.²⁶) Finally, there is the observer's sympathy with the recipient's gratitude to the agent. This

²⁶ Thanks to John McHugh for alerting me to some of the difficulties indicated here.

sympathetically-acquired feeling is an actual feeling of gratitude *to* the agent, so *she* is clearly its intentional object. Thus Smith can say that the feeling of merit takes the agent as its object in just the way that any feeling of gratitude takes someone as its object.

The sense of demerit has all the analogous parts, including a sense of propriety toward the recipient's distress, a sense of impropriety caused by the discord between the observer's feeling and that of the agent, and ultimately, the sympathetically-acquired resentment toward the agent, which takes the agent as its intentional object in just the way (whatever that is) that anyone who is resented is the object of that emotion. Given the important role of resentment in constituting the feeling of demerit, it is fairly clear how this moral sentiment (the feeling of demerit) focuses on the third party in the story, even though it is caused in part by sympathetically-acquired feelings of distress that had no intentional object or a different one. The sense of demerit has this intentional object because its crucial component, what mainly distinguishes it from the sense of propriety (alone), is resentment, which by its nature has just this intentional object. Thus the gratitude or resentment felt by the recipient already has the agent as its intentional object. The directedness of the observer's feeling of merit and demerit, and its specific target, are determined by the nature of the recipient's sentiment that the observer comes to share.

Now, gratitude and resentment are not merely pleasant and unpleasant feelings that cause one to think of a person. They are feelings that respond to the proper or improper motives of another, feelings that thus can only be directed to human beings who harbor emotions or desires toward the recipient that are, or are not, capable of being mirrored by an impartial observer. If Andrew has harmed another by his deceitful behavior (spurred, let us

suppose, by greed), an observer cannot resent Andrew's knit cap for this, nor his innocent parents (who, let us suppose, are neither greedy nor deceitful). That is not how the emotion of resentment works.

It is worth noting how much P. F. Strawson's account of moral indignation and moral disapprobation parallels Smith's account of the sense of demerit. To be morally indignant on behalf of another or to feel moral disapproval of someone's behavior is, for Strawson, to "experience[] the vicarious analogue of resentment"; and "it is this impersonal or vicarious character of the attitude... which entitles it to the qualification 'moral.'" For Strawson (1962) the reactive attitude of resentment, when "generalized," becomes moral condemnation. Many today find this Strawsonian account of moral sentiments quite attractive. But it was Smith who developed it as an account of some of the crucial moral sentiments within a full-blown moral sentimentalism.

So we should look more closely at how Smith understands the relation between gratitude and resentment and their intentional objects.

To do this we must consider Smith's account of what sort of sentiments gratitude and resentment are. Gratitude is "the sentiment which most immediately and directly prompts us to reward" someone, and resentment is "that which most immediately and directly prompts us to punish" someone (TMS II.i.1.2, p. 68). Further, gratitude prompts me to reward someone's benefactor myself, whether or not she benefited me, and to the same degree to which the benefactor actually benefited the recipient. Resentment makes me desire to punish the one who did the harm, regardless of who received it (myself or another); and in feeling resentment I even desire to mete out this punishment myself, in exact proportion to the harm the agent

caused, and by inflicting the same type of harm on her, in order to make the agent regret her harmful act (TMS II.i.1.6, p. 69). So gratitude and resentment, according to Smith, are desires to do something to or for the person who inflicted the harm or provided the benefit. And we have seen that our moral evaluation of the merit or demerit of someone's motive or action in part *is* a feeling of gratitude or resentment, for Smith.²⁷ So the moral sentiments of merit and demerit are directed to their objects in exactly the way that desires (to reward and to punish) are directed to their objects.²⁸

c. More About Smith's Sense of Merit and Demerit: some Difficulties for
Resentment

Smith's appeal to gratitude and resentment has an advantage: it yields a far more specific and more satisfying account of how the moral sentiment is directed toward its object, the person we evaluate, than Hume's theory does. The sense of demerit I feel may, along the way, cause me to think of Andrew's knit cap or his innocent parents, but from Smith's account it does not follow that I morally blame Andrew's cap or parents, because I do not resent them.

²⁷ By contrast, Hume does not use the passions of gratitude or resentment to explain the causation or composition of the moral sentiments at all.

²⁸ Smith mentions a further aspect of the passions of gratitude and resentment. Gratitude, as a desire to *reward*, is not only a desire to make our benefactor feel pleasure and to make him know that he feels it on account of his beneficence to us, but a desire to confirm in him his high opinion of us as worthy of his ministrations. Resentment, as a desire to punish, is not only a desire to make one who harmed us suffer and know he suffers on account of his mistreatment of us, but a desire to correct his contemptuous opinion of us (TMS II.iii.1.4-5, pp. 95-6). I take these to be Smith's analyses of the desires to reward and punish. The wording here is clearly the language of desires and goals.

At least we have an account of the intentional objects of moral disapproval that is more discriminating than one based on mere association and causal regularity.

But notice what has actually happened, given Smith's understanding of what gratitude and resentment themselves are. Smith achieves greater precision in understanding the relation between certain moral sentiments and their intentional objects by construing the sense of merit and demerit as desires, indeed desires for very specific goals. The desire in question is not, as in Hobbes, just a desire for something or other that then leads us to call it good. But the moral sentiment of merit or demerit is a passion with the structure of a desire: it is an urge to do something that the observer aims at. Smith does not say whether he understands desires and aversions as Hume does, as nearly all aimed at attaining some prospective pleasure for the person who feels them or avoiding some uneasiness. But whether he follows this Humean precept or not, Smith thinks of gratitude as an impulse to reward the agent, and he thinks of resentment as an impulse to punish the agent. While there is more to the sense of merit and demerit than gratitude and resentment, they are, in part, these desires. Thus Smith's use of gratitude and resentment to compose the sense of merit and demerit, given his account of what gratitude and resentment are, yields a better account of their intentionality at the cost of turning them into desires. And this conflicts with something I have argued is an important feature of our moral experience, something Hume does capture by making the moral sentiments indirect sentiments rather than desires. The sense of demerit becomes, for Smith, an urge to achieve a goal: punishment. Whereas moral judgments, while they are often linked with such urges, are not identical with them. Moral evaluations are not essentially goal-seeking.

I should say more in support of this claim. Of course, resentment often accompanies the judgment that an action was wrong. But it is less plausible to say that it constitutes or forms a necessary part of that judgment, at least if we understand resentment as Smith does, because there are many occasions when we judge an action wrong, and judge it to be so because it inflicted foreseeable and unjustified harm, and yet we do not desire to punish the agent nor even desire that she be punished. This can be because the agent “has already suffered enough,” as people sometimes say of parents whose young children were accidentally killed as a result of the parents’ own negligence. Or it can be because we wish to do no harm to anyone. Or it may be because the whole matter seems not to be our responsibility or is too remote to evoke such a desire in us. (I do not wish to punish Rasputin for his evil deeds, nor do I wish that he had been punished (perhaps I don’t care), but nonetheless I consider them evil.) Or there can be moral disapproval without a desire to punish because the victim has forgiven the perpetrator. When I forgive your wrongful act toward me, I do not cease to judge that you did wrong; nor would a neutral observer do so. Your motive at the time of action still had demerit. But I do not wish to punish you. Indeed, one way to describe what I do in forgiving your action is that I foreswear resentment. But if I bear you no resentment, by Smith’s account I fail to judge your action to have been wrong.

A mordant remark of Heinrich Heine’s comes to mind: “Yes, we ought to forgive our enemies, but not until they are hanged” (1873, p. 83). If we follow his advice, our resentment will be gratified first, before we forgive and so cease to resent them. If we disobey the advice and forgive those who wronged us *before* they are punished, then we will no longer want to punish them – and where is the satisfaction in that? But if we do forgive them before they are

punished, and so we cease wanting to punish them, we still judge them to have done wrong, which is a problem for Smith's account.

This shows, too, why we cannot construe Smith as thinking of the sense of demerit as a desire to punish that is defeasible by various conditions (such as forgiving). As a sentimentalist, Smith thinks that the (vicarious) feeling of resentment just *is* the judgment that the action was wrong, or at least is a necessary part of it. Even when it is not advisable actually to punish the agent because of a defeating condition (it would create social unrest, it would damage a relationship, and so on), in order to judge the action as having demerit at all we must feel this desire to punish. If we cease wanting to punish, we cease thinking the action wrong. That simply seems incorrect.

IV. Summing Up

Hume's moral sentiments are affective states, are pleasant or uneasy, and have persons as their intentional objects; but, to Hume's credit, in and of themselves they are not urges. Because of their emotional quality they can play an indirect role in moving us to act and to feel in a variety of ways without themselves being impulses, which seems consistent with the variety of roles that ethical evaluations play in our affective and practical lives. Hume notes that moral judgment shifts our attention from recipient to agent, which is also a strength of his account. But unfortunately his account of what it is for a moral sentiment to be about or of someone, a feature he needs for a successful moral sentimentalism that fully accounts for our moral judgments as manifestations of our emotions, is crude and inaccurate. The problem seems to be the result of excessive parsimony: Hume starts with such a small set of basic

psychological states and relations between them that he has difficulty accounting for the complexity of some of them, including our ethical responses.

Smith rather easily narrows down a specific relation between one type of moral sentiment and its intentional object by appealing to two very familiar reactions to receiving good or harm at another's hands: gratitude and resentment. But in attempting to describe what moral evaluation is like in and of itself, Smith goes wrong in making this type of moral sentiment essentially goal-directed. Thinking someone morally bad and wishing to punish can diverge, but on his account they cannot. The account of intentionality implicit in Smith's psychology of the moral sentiments is better, but the account Smith must give of what our moral sentiments are like is flawed as a result.

These two desiderata, a convincing account of the relation between a moral sentiment and its intentional object on the one hand and a description of the moral sentiment that tracks our patterns of moral judgment on the other, seem to be in tension.

For those attracted to the Smithian, and/or to the Strawsonian, account of moral disapproval, on which that disapproval is a generalized form of resentment, one possible strategy would be to give a different and more plausible account of what resentment is that does not construe it as a goal-seeking impulse to punish, so that the Humean idea of moral disapproval as an assessing sentiment rather than an urge might be accommodated. I don't have such an account to offer here, but encourage those who are attracted to such a view to develop it. Alternatively, a moral sentimentalist might abandon the appeal to generalized or vicarious resentment, and seek a different emotional basis of moral disapproval. With either

approach, there is much careful work still to be done to meet the two desiderata simultaneously and generate a satisfying moral sentimentalism.

These two eighteenth-century figures, Hume and Smith, did more to articulate the details of how the moral sentiments work than most present-day moral sentimentalists. Yet, if I am right, they both failed to fulfill a pair of desiderata, though in opposite ways. I have certainly not shown that these two purposes cannot be achieved simultaneously by any sentimentalist theory of morality. But we also have no guarantee that they can.²⁹

²⁹ Thanks to participants in the conference “The nature and origin of morality: Adam Smith’s response to David Hume’s views on moral matters” in Oslo, Norway, organized by Christel Fricke and Lilli Alanen under the auspices of the Centre for the Study of Mind in Nature, the University of Oslo, August 2016; to the audience at the Symposium: Moral Sentimentalism and Its Foundations, California State University, Fullerton, March 2017, and the audience at the 44th Hume Society Conference at Brown University, July 2017, and my commentator Lauren Kopajtic, for fertile discussion of various early and rather different drafts of this paper. Thanks to Bradley Armour-Garb for invaluable discussion at a transition point and to John McHugh for very insightful written comments.

References

- Anslie, Donald. (1999). "Scepticism about Persons in Book II of Hume's *Treatise*," *Journal of the History of Philosophy* 37, pp. 469-92
- Árdal, P. (1998). *Passion and Value in Hume's Treatise*, Edinburgh: Edinburgh University Press
- Cohon, R. (2008a). "Hume's Indirect Passions" in *A Companion to Hume*, Elizabeth S. Radcliffe Ed., Oxford, U.K.: Blackwell Publishing
- Cohon, R. (2008b). *Hume's Morality: Feeling and Fabrication*, Oxford, UK: Oxford University Press
- Cohon, R. (2010). "Hume's Moral Sentiments as Motives," *Hume Studies* 36;2 (Nov.), 193-213
- Fleischacker, S. (2012). "Sympathy in Hume and Smith: A Contrast, Critique, and Reconstruction," in *Intersubjectivity and Objectivity in Adam Smith and Edmund Husserl*, ed. Christel Fricke and Dagfinn Føllesdal, pp. 273-311, Frankfurt, Germany: Ontos Verlag
- Heine, H. (1873). *Scintillations from the Prose Works of Heinrich Heine*, Simon Adler Stern, Trans., New York, NY: Henry Holt and Co.
- Hobbes, T. (1651*1996 rpt 1999?) *Leviathan*, Richard Tuck, Ed., Cambridge, U.K.: Cambridge University Press
- Hume, D. (1992, 1995) *Dissertation on the Passions* (1757), in *Four Dissertations and Essays on Suicide and the Immortality of the Soul*, South Bend, IN: St. Augustine's Press
- Hume, D. (1751/2004) *An Enquiry concerning the Principles of Morals* (1751), ed. Tom L. Beauchamp (Oxford University Press, 1988 rpt. 2004)
- Hume, D. (2000) *A Treatise of Human Nature* (first published 1739-40), David Fate Norton and Mary J. Norton (Ed.s), New York: Oxford University Press
- Hutcheson, F. (1724/1994) *An Enquiry concerning the Original of Our Ideas of Virtue or Moral Good* 1725/1994, rpt. from 4th edition (1738) in *Philosophical Writings*, ed. R. S. Downie,

Everyman Library (Orion Publishing Group, London, and Charles E. Tuttle Co., Inc.,
Rutland, VT

Lewis, D. (1989). "Dispositional Theories of Value," *Aristotelian Society Supplement* 63 (113-37)

Loeb, Louis. (1977). "Hume's Moral Sentiments and the Structure of the *Treatise*," *Journal of the History of Philosophy* 15, pp. 395-403

Moore, G. E. (1903, rpt. 1976). *Principia Ethica*, Cambridge, UK: Cambridge University Press,

Radcliffe, E. S. (2018). *Hume, Passion, and Action*, Oxford, UK: Oxford University Press

Smith, A. (1790 /1982) *The Theory of Moral Sentiments*, D. D. Raphael and A. L. Macfie, Ed.s,

Indianapolis: Liberty Press,

Spinoza, Baruch (1677). *Ethics*. G. H. R. Parkinson (Ed. and Trans.) New York: Oxford University Press, 2000.

Strawson, P. F. (1962). "Freedom and Resentment," *Proceedings of the British Academy* 48, pp. 187-211