Spring 2016

What motive to virtue? Early modern empirical naturalist theories of moral obligation

Brady John Hoback
University of Iowa

WHAT MOTIVE TO VIRTUE? EARLY MODERN EMPIRICAL NATURALIST THEORIES OF MORAL OBLIGATION

by

Brady John Hoback

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Philosophy
in the Graduate College of
The University of Iowa

May 2016

Thesis Supervisor: Professor Diane Jeske

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Brady John Hoback

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Philosophy at the May 2016 graduation.

Thesis Committee: _____
Diane Jeske, Thesis Supervisor

_____
Asha Bhandary

_____
David Cunning

_____
Richard Fumerton

_____
Ali Hasan

To the loving memory of Jake and Shawna

ACKNOWLEDGMENTS

I must extend my sincere gratitude to the many people who made this dissertation possible. I am deeply grateful for my wonderful, hard-working, and inspiring family; my parents Giff and Judi Hoback, my sister and brother-in-law Tracy and Kent Bozarth, my sister Lindsay Hoback, and my grandmother Paula Geary have all been unwavering in their love and their support. I love you all dearly. Thank you to Mandy Yoches for a rewarding 16 years of friendship. Thank you to Katie Biesendorfer and Phil Ricks for being such great friends, roommates, and interlocutors. Many thanks to Joe Ratcliff for making these last months of graduate school so much brighter. Thank you to LaShay Peterson, Aaron Gray, Sam Trammell, and Cheryl Hetherington for all your help along the way.

While I am indebted to many of the faculty in the philosophy department, I owe a special thanks to Diane Jeske for introducing me to the central issues and figures in the history of ethics, and for her incredible patience with me as I struggled to understand them. Thank you to David Cunning for serving as director of my comprehensive exam committee. I would also like to thank Richard Fumerton and Ali Hasan for the many helpful discussions about ethics that we had while I served as a TA for their Introduction to Ethics courses. Finally, thank you to Asha Bhandary for her outstanding classes in ethics, feminism, and political philosophy.

I would like to thank the philosophy graduate students at the University of Iowa, past and present, for their encouragement and for the role they played in helping me grow as a philosopher. While I thank you all for helping to foster the close-knit, supportive, and productive environment that is characteristic of our department, I am especially grateful to Kris Phillips, Sam Taylor, Greg Stoutenburg, Chris Dyer, Dave Redmond, Ian O'Loughlin, Abe Graber, Pete LeGrant, Seth Jones, and Landon Elkind. Thank you for your guidance and support, for consistently beating the intellectual snot out of me, and for helping to make graduate school one of the most rewarding experiences of my life.

ABSTRACT

In this dissertation, I argue for a set of interpretations regarding the relationship between moral obligation and reasons for acting in the theories of Hobbes, Hutcheson, and Hume. Several commentators have noted affinities between these naturalist moral theories and contemporary ethical internalism. I argue that attempts to locate internalist theses in these figures are not entirely successful in any clear way. I follow Stephen Darwall's suggestion that addressing the question "why be moral" is one of the fundamental problems of modern moral philosophy. Since, as some have argued, there is a tension between accepting internalism and providing an adequate response to the "why be moral" question, I argue that each figure maintains a distinctive response to this question given the sort of internalism, if any, he would accept. In the introduction, I provide the key distinctions that arise from contemporary discussions of ethical internalism, and I motivate my project of looking for insight into the relationship between internalism and amoralism in the British Moralists.

Chapters 1 and 2 focus on the moral theory Hobbes who, I argue, would accept a version of existence internalism because he holds that there is a necessary connection between one's being contractually obligated and one's being in certain rationally motivating states. I then present the fool's objection as an objection to the assumption of a relevant similarity between divine obligation and contractual obligation. I argue that, irrespective of this dissimilarity, the fool has some rational motive to keep his covenants in virtue of the fact that making covenants changes one's decision situation in such a way that it becomes reasonable to treat covenants as if they imposed categorical constraints on behavior. I claim that Hobbes's response to the fool is, at least in part, that the fool fails to understand what moral obligations are.

In the remainder of the dissertation I turn my attention to two classical sentimentalist moral theories. I examine the theories of Hutcheson and Hume because it is not clear what resources moral sentimentalism has available to it in order to address questions about the reasonableness of moral action. In chapters 3 and 4, I develop an interpretation of Hutcheson

who, because he distinguishes between exciting and justifying reasons, is able to say there is some non-derivative sense in which moral actions are reasonable. I argue that he develops a theory whereby moral obligation is to be understood in terms of the non-motivating states of approval of moral spectators, and I do not think, contrary to Darwall, that there is anything puzzling about his doing so. I argue that Hutcheson does not accept a version of motive internalism, but that he shares much in common with internalist views: he claims that there is a very strong, if contingent, connection between our states of approval and our motivational states. I offer an explanation of how Hutcheson could respond to the amoralist, which holds that we ought to be moral because, in part, we all already have the motives for and the interests in doing the sorts of things of which moral spectators approve.

In chapters 5 and 6, I turn my attention to Hume who, because he makes no distinction between motivating and justifying reasons, does not seem to have anything to say about the non-derivative reasonableness of moral action. I argue that a textually grounded interpretation of Hume's theory of the passions provides us with more reason to favor an (appraiser motive) internalist reading over an externalist reading of his moral theory. Much of my argument depends on an interpretation of Hume's claim that it is possible for agents to be moved to act from a sense of duty alone. When we ask what Hume can say to the question "why be moral," some of the options that Hutcheson pursues are initially open to him. However, I argue that Hume thinks philosophical theorizing must give way to the operations of psychological mechanisms that are causally responsible for inspiring agents to act morally by giving rise in them to particular kinds of affections.

I conclude with some general remarks about the problems surrounding Darwall's interpretation of Hume's theory of justice, and use this discussion to lend further support to the claim that the actual theories of Hobbes, Hutcheson, and Hume do not neatly fit into the taxonomies that Darwall seems to think they do.

PUBLIC ABSTRACT

We are all familiar with the difficulty that can attend fulfilling our moral obligations, especially when doing so involves giving up something that we desire or requires us to do something that makes us uncomfortable. The intellectual upheaval of 17th and 18th Century European thought sought an end to the worldview whereby God serves to provide purpose and order to everything in the universe; and this meant that it was no longer satisfactory to say that we should be moral because of God's rewards and sanctions.

My research focuses on one tradition that emerges from this intellectual climate, called empirical naturalism. This tradition includes Thomas Hobbes, Francis Hutcheson, and David Hume. It holds that morality can be explained in entirely naturalistic terms and that human reason is only capable of discovering truth and falsity. It is a philosophical tradition that, in various ways, grounds our reasons for being moral in certain psychological traits of human nature.

Some contemporary philosophers see the empirical naturalists as holding a theory known as ethical internalism. This theory says that there is a necessary connection between moral obligation and an agent's reasons for acting. There is a tension between accepting internalism and asking "why be moral?" My dissertation resolves this tension in Hobbes, Hutcheson, and Hume. I believe doing so gives us a better understanding of the nature of the difficulty we sometimes face in doing the right thing.

TABLE OF CONTENTS

INTRODUCTION
ETHICAL INTERNALISM AND THE QUESTION "WHY BE MORAL" IN THE BRITISH MORALISTS

Introduction

In this dissertation, I argue for a set of interpretations regarding the relationship between moral obligation and reasons for acting in the theories of Hobbes, Hutcheson, and Hume. Since all of these naturalistic theories are grounded in claims about human nature, I will also offer a set of interpretations regarding their different accounts of moral psychology. It is my aim to present these views on their own terms. Hobbes, Hutcheson, and Hume are all systematic thinkers and it is hard to say anything about particular elements of their accounts without succumbing to the temptation to try to say something about all of the other elements of their views. The temptation is that much harder to resist when one is dealing with figures who say so much that is interesting, controversial, puzzling, exciting, and at times even beautiful. The reader will be reassured that I have resisted this temptation to the extent that I think wise. There are certain systematic considerations that need to be addressed in order to provide an adequate interpretation of the question of moral obligation and its relation to reasons for acting; questions about the meaning of moral terms, about the nature of sensation and perception, the operation of passions and affection, and the nature of will are some of the important ones with which I deal.

Many philosophers have thought that the concept of 'obligation' is necessarily linked with reasons or motives for action. This is, very broadly, ethical internalism, and in this introduction I will specify more clearly what such claims amount to. I mention internalism here because one of my overall aims is to follow Stephen Darwall's provocative suggestion that the roots of this contemporary metaethical thesis can be found in the British Moralists. In what sense, if any, can we think of Hobbes, Hutcheson, and Hume as developing internalist moral theories? Given Hume's famous remarks that morality has an influence on our passions and our actions, his internalist commitments might be taken as settled. However, as we will see in

chapter 6, it is by no means uncontroversial what Hume means by these remarks. Nevertheless, the reader might initially find some plausibility to the suggestion that Hume would accept some version of internalism. It is well known that Hutcheson was a major influence on Hume's views of the nature of morality. These affinities might lead one to suppose that Hutcheson, too, would have accepted a version of internalism. It is also well known that both Hume and Hutcheson, in good post-Hobbesian British Moralist fashion, reject significant elements of, what they took to be, Hobbes's account (though Hutcheson does so with quite a bit more fervor). So, one may be inclined to think that Hobbes denies a necessary connection between obligation and motivation. These initial impressions, however, will not stand to further scrutiny. In fact, I argue that it is Hobbes's view that is most amenable to being read as a certain version of internalism (I develop this argument in chapter 2). While I do argue that there are some considerations favoring an internalist reading of Hume, I argue that they are by no means decisive (I turn to these consideration in chapter 6). Hutcheson, however, cannot be read as an internalist in any strict sense (see chapter 4).

One of the important results of my discussion about internalism in these three figures is that it begins to put pressure on what Darwall understands as a distinct tradition of internalist theories that arose in early modern British moral philosophy. Hobbes, Hutcheson, and Hume all fall within the tradition that Darwall calls *empirical naturalism*. I will have more to say about this tradition in a later section of this introduction. I mention it here because one of the major problems that arises for Darwall's discussion is how exactly to understand empirical naturalism as an *internalist* tradition. In chapter 7, I offer some considerations in favor of the view that Darwall's attempt to locate such a tradition in Hobbes, Hutcheson, and Hume is unsuccessful.

Nevertheless, I do think that attempting to locate internalist commitments in the British moralists, generally, is an interesting and worthwhile project. One of the concerns of contemporary philosophers who accept internalism is with justifying morality, or accounting for

an obligation to be moral, or providing an answer to the question "why be moral?"[1] It is not entirely clear to me how internalism works as a justification for morality, since the internalist seems to hold the view that morality stands in no further need of justification. Moreover, as we will see in this introduction, the idea that the internalist can justify morality is contentious. I do not intend to give a full examination of these issues here. What I want to signal is that contemporary discussions of internalism arise right alongside discussions about the question "why be moral?" This is important for my project because I follow Darwall in thinking that the "why be moral" question was an important one for early modern British moral philosophy. I accept Darwall's claim that this is the "modern moral problem." Since Darwall argues that the roots of internalism are to be found in the British Moralists, it seems to me interesting and important to figure out how exactly these commitments are supposed to be related to the "why be moral" question. How, if at all, is the "internal ought" supposed to provide a "justification" of morality?

These comments are intended as introductory remarks and they are admittedly vague and underdeveloped. I will not ask more of the reader's patience with these sorts of remarks. In the next section, I distinguish the various kinds of internalist theories that I discuss throughout the dissertation. This will involve laying bare a set of conceptual resources that, while not entirely distinct from the resources with which Hobbes, Hutcheson, and Hume were working, is not developed explicitly in their own terms. My dissertation is an attempt to specify the extent to which any of the empirical naturalists, taken on their own terms, would have accepted *these* theories. In the subsequent section, I provide a brief discussion of the two "internalist" traditions that Darwall identifies. In the final section, I discuss the modern moral problem in more detail.

---

[1] Coleman 1992, 331. Nagel argues that one of the deficiencies of externalism is that it allows "someone who has acknowledged that he should do something and has seen *why* it is the case that he should do it, to ask if he has any reason for doing it." Quoted in Coleman ibid.

Some Preliminary Distinctions

Internalism is typically taken to be the view that there is a necessary connection between one's making a moral judgment and one's having a reason to do as she judges. This statement of the view is too ambiguous to be of any help in understanding or evaluating the internalist thesis, and below I will specify the different ways in which this claim has been understood. Before laying out the various versions of internalism that will concern me, however, it will be useful to say something about the important concepts that philosophers have used to distinguish the many varieties of internalism.

Reasons and Motives

First, one's having a reason to do as she judges is ambiguous between one's having an *explanatory*, *motivating,* or *exciting* reason, and one's having a *justificatory* or *good* reason for acting. For example, imagine that a police officer stops a young man for driving 55 MPH in a residential area. In the course of their conversation, she asks him, "Why were you going so fast?" The young man, who figures that honesty is the best strategy in this situation, tells her that he was running a bit late to his second date with a young woman who previously said she highly values punctuality. Presumably, both the boy's desire to impress the girl, and his belief that speeding is required to appear punctual, explain his behavior. The officer now knows about some of the beliefs and desires that move him to act. However, what we don't get, and what from the officer's perspective might be really tough to get, is anything like a good reason for his behavior. The beliefs and desires (the motives) of the young man do not necessarily justify his speeding.[2]

One might, however, wonder about what exactly counts as a *good* reason. What is a good reason if it is not merely a matter of one's desires or motivations? There is the further, related question whether the justification here is to be understood in an objective sense, i.e. the agent is actually justified, or is to be understood merely in a subjective sense, i.e. the agent takes himself

---

[2] For a similar example, see, Brink 1989, 40. Both my own and Brink's examples presuppose a Humean theory of motivation, *viz.* that motivation requires both beliefs and desires.

to be justified. Also, one might wonder if the existence of a good reason is entirely independent from the existence of one's actual or dispositional motives, or if there is some important connection between them.[3] These are interesting and important concerns, and as we will see in chapters 5 and 6 Hume seems to provide some controversial responses to them. All I hope to do here is to point to the basic distinction between exciting and justifying reasons: exciting reasons can explain behavior without justifying it, and justifying reasons give good reasons for behavior (perhaps independently of one's motivation, though I leave that question open for now).[4]

Moral Concepts: Broad and Narrow Interpretations

Secondly, since the internalist is asserting the existence of a necessary connection between reasons for action and moral judgment, we need to be clear on what this moral judgment consists in. There are two concerns relevant here. The first has to do with the term *moral* in 'moral judgment.' One might take a broad interpretation, in which case the internalist will assert the existence of a necessary connection between *any* moral judgment and reasons for action. Alternately, one might take a narrow interpretation, and specify that the necessary connection doesn't hold between just any moral judgment and reasons for action, but that it holds between judgments involving a particular moral concept, or set of such concepts, and reasons for action.[5] In chapter 2, I argue that Hobbes adopts a broad conception. The question is not so clear with

---

[3] Falk and Smith both seem to accept the view that one has a justificatory reason for action just in case one also has some actual or dispositional motive to act. See Falk, 1948. For a discussion of Falk, see Darwall 1992, 155-156. See also Smith 1994; reprinted in Shafer-Landau 2007, 231-242.

[4] The distinction here is often made in terms of *exciting* and *justifying* reasons. As Darwall notes, the distinction between justifying and exciting reasons is typically attributed to Hutcheson. However, because Hutcheson has a particular conception of the nature of a justifying reason, as a motive which engenders approbation, it isn't clear that Hutcheson's distinction maps perfectly onto this, broader, distinction. See Darwall 1992, 170.

[5] I have chosen to talk about the concept of moral obligation here. I will later refer to moral obligation as a moral property, or to one's being under an obligation as a moral fact. I recognize that moral concepts, properties, and facts are distinct, and I do not want to suggest that they are interchangeable here. For the purposes of this introduction, I do not think specifying this distinction will help get a clearer picture of the internalist's views. I will come back to this distinction when doing so will yield a fruitful discussion.

respect to Hutcheson and Hume (chapter 4 and 6). For the purposes of this introduction, I will stipulate that the term *moral* is to be given a narrow interpretation. The moral concept that will concern me here is that of 'moral obligation,' or 'being such that one ought (morally) to do it.'[6]

Moral Judgments: Sincerity and Truth

The second relevant concern arises from two different ways in which one might evaluate moral judgments. For example, let's stipulate that Finn is a particularly flakey and unreliable person. When you make plans with him he always backs out at the last minute, you can expect him to never arrive on time to appointments, etc. He isn't a particularly absent-minded person, and he takes great pains to remind himself of his commitments (he keeps a calendar, a planner, writes sticky notes, and sets alerts on his phone). Imagine further that in the course of a rather serious and reflective conversation with Finn, you hear him say, "One ought to keep his engagements." One natural reaction to hearing this utterance is to question the sincerity of Finn's holding it. We might, alternately, think that Finn lacks even a modicum of self-awareness, and so think that he can sincerely accept his claim about keeping one's engagements while nevertheless not keeping his own engagements. Let us, then, stipulate further that you have good reason to think that Finn is a particularly self-aware individual. Given this, we might then just deny that he really believes what he says when he says that one should keep his engagements. To the extent that he has made a moral judgment, he has failed to make an authentic or sincere moral judgment. He's given voice to a platitude that he himself does not accept, so his moral judgment is not *genuine*. Nevertheless, Finn might be correct in claiming that one ought to keep one's engagements. Even if we question his sincerity in accepting the moral judgment, we can still recognize that he is, or might be, saying something true. So, we can evaluate moral judgments in terms of the sincerity with which they're held, or in terms of their truth-value.

---

[6] I mean for these two concepts to be interchangeable here. So, by 'moral obligation' I do not mean to pick out *prima facie* obligations. As I understand it here, if one is morally obligated to Φ, then Φ is what one ought (morally) to do. At least, this is how I will use 'moral obligation' throughout the course of the introduction.

Necessary Connection: Constitutive and Causal

The internalist is asserting the existence of a necessary connection between moral judgments (either those sincerely held, those that are true, or those that are both sincerely held and true) and reasons for action (either explanatory or justifying). So, the internalist will need to specify the nature of this necessary connection. There are two important ways in which she can do this. First, the internalist can say that the necessary connection obtains in virtue of some constitutive relationship between moral judgment and reasons for action. For example, one might hold that all there is to making a sincere moral judgment is having certain reasons for action. Similarly, one might hold that the truth of moral judgments is nothing over and above the fact that the agent making them has certain reasons for action. Alternately, one might hold instead that there is a necessary causal connection between making a moral judgment and one's reasons for action. This sort of a view would not hold that moral judgments (either sincere or true) were made up of reasons for action. Rather, they would hold that moral judgments, when sincere or true, have necessary causal consequences on our reasons for action.[7]

Reason Internalism and Motive Internalism

With these preliminary distinctions, we can begin to spell out in more detail the different versions of internalism that will be dealt with throughout. First, we can distinguish between reason internalism and motive internalism as follows:

> Reason internalism: necessarily, if an agent, S, is under an obligation to Φ (alternately, if she genuinely judges that she is obligated to Φ), then S has *some* justificatory reason to Φ.

[7] There are two implications from this section that I wish to avoid. The first is the suggestion that all causal relationships are necessary relationships. The internalist need not be committed to this view. Instead, the internalist who adopts a causal theory may insist only that there is at least one necessary causal relationship. Secondly, the internalist may hold that there is a necessary connection between moral judgment and reasons for action that is neither constitutive nor causal. The details of such a position need not concern us in this introduction, since the two most important ways of understanding the necessary connection are either constitutive or causal.

> Motive internalism: necessarily, if an agent, S, is under an obligation to Φ (alternately, if she genuinely judges that she is obligated to Φ), then S has *some* motive to Φ.

Let's say, for example, that Jane is obligated to tell the truth. According to reason internalism, it is necessarily the case that, if she is so obligated, Jane is also to some extent justified in telling the truth.[8] Jane has a good reason to tell the truth. Now, depending on one's analysis of having a good or justificatory reason, it might not follow from Jane's being obligated to tell the truth that she also has some motive to tell the truth. The existence of a justificatory reason can, but need not, entail the existence of any particular motives of the agent. This is not the case for the motive internalist. Necessarily, if Jane is obligated to tell the truth, then, for the motive internalist, Jane also has some motive to tell the truth. But the mere fact of Jane's having a motive to tell the truth does not *ipso facto* show that Jane is justified in telling the truth. So, if she were to tell the truth, her behavior could be explained in terms of her motives, but we might still think she had no good reason to tell the truth.

In what follows, I specify the versions of internalism offered by Darwall. Since Darwall focuses exclusively on motive internalism, I will continue on in this introduction by only spelling out different versions of motive internalism.[9] It should be noted, however, that for every version of motive internalism that I specify, there is an analogous version of reason internalism.

Appraiser and Existence Internalism

We can distinguish between two broad kinds of internalism by focusing on the distinct ways in which moral judgments can be evaluated. Again, moral judgments can be evaluated in terms of their being sincerely accepted by those who make them, or they can be evaluated for their truth-value.[10] This gives us the following views:

---

[8] All I mean here by saying that she is to some extent justified is that she has some justificatory or good reason for telling the truth.

[9] It will become clear that, because Darwall often talks about a *rational* motive, the motivating/justifying distinction is not of chief importance for him.

[10] As mentioned above, this disjunction is to be understood inclusively.

> Appraiser internalism: necessarily, if an agent, S, genuinely judges that she ought to Φ, then S has some motive to Φ.
>
> Existence internalism: necessarily, if an agent, S, is obligated to Φ, then, under certain circumstances, S will have some motivation to Φ.

For reasons that will shortly become clear, I prefer Brink's nomenclature, which is given above, for the first view to Darwall's label of 'judgment internalism'.[11] To illustrate the position it will be helpful to recall the case of Finn the flake that was discussed above. According to appraiser internalism, it is necessarily the case that if Finn genuinely judges that he ought to keep his engagements, then Finn has some motive to keep his engagements. Remember that we supposed that Finn was particularly self-aware, and was remarkably bad at keeping his engagements. If it's true that Finn lacks *any* motive to keep his engagements, then it cannot be the case, according to appraiser internalism, that Finn genuinely judges that he ought to keep his engagements.[12]

As Darwall points out, the appraiser internalist position seems to be held by many twentieth-century noncognitivists. For example, Stevenson maintains that it is part of the "vital" sense of *good* that it has "so to speak a magnetism. A person who recognizes X to be 'good' must *ipso facto* acquire a stronger tendency to act in its favour than he otherwise would have had."[13] A moral judgment, for Stevenson, is a linguistic expression of, in part, the agent's strong interests in, or pro-attitudes towards, that which she "judges" to be good. Since these interests are

---

[11] Strictly speaking, my formulation of appraiser internalism is distinct from Brink's. Here is Brink's formulation of the view: "*Appraiser internalism* claims it is in virtue of the concept of morality that *moral belief* or *moral judgment* provides the appraiser with motivation or reasons for action" (Brink 1989, 40). The difference between Brink's characterization and my own, then, is simply that I am, for the purposes of this introduction, claiming that appraiser internalism is a version of motive internalism alone. Note, however, that this is how Darwall characterizes judgment internalism: "it is a necessary condition of a sincere or genuine ethical or normative utterance, thought, or conviction…that one would, under appropriate conditions, have some motivation to x." (Darwall 1995, 9; 1992 155; 160-162).

[12] Of course, one might understand the situation differently. It might not be the case that Finn lacks *any* motive to keep his engagements. The flakey person might be motivated to a very small extent to keep his engagements. He is flakey insofar as other, stronger motives are always present when it comes time to act. For the purposes of illustration, however, I will suppose that Finn lacks any motive to keep his engagements whatsoever.

[13] Stevenson (1937) 1970, 256. Stevenson seems to take 'good' as a stand in for any moral concept. If so, then we have an example of an internalism that is not merely about the concept of moral obligation, but is instead a broad version of internalism.

expressed in the moral judgment, Stevenson thinks his view accounts for the internalist requirement in a straightforward way. The agent already has an interest in that which she "judges" to be good. These interests, in the moral case, are understood in terms of one's experiencing "a rich feeling of security when it [the object of approval] prospers," and being "indignant or 'shocked' when it does not."[14] It is plausible to suppose that feelings of security and indignation are motivating sentiments. Since moral judgments are expressions of these conative, motivating states, it cannot be the case that one makes a moral judgment without thereby having some motive.[15]

I find it odd that Darwall refers to positions held by twentieth-century metaethical noncognitivists, like Stevenson and Hare, as varieties of *judgment* internalism. This is because, strictly speaking, it's not clear that Stevenson believes there are any such things as moral judgments *per se*. If we understand judgments to be cognitive descriptions that purport to provide truth-evaluable information about the world, then Stevenson certainly cannot think that there are moral judgments. It is for this reason that I prefer Brink's label for this version of internalism, and will refer to it as *appraiser internalism* in what follows.

Before illustrating the existence internalist's position, it will be helpful to first specify the qualification "under certain circumstances" that I used in the formulation of the thesis above. While there may be many considerations that will fall under these circumstances, for the purposes of this introduction I will only briefly discuss one such consideration. Motive internalism would be a problematic view indeed if it held that moral properties or states of affairs, such as moral obligation or one's being morally obligated, were necessarily connected to an agent's motives *independently* of an agent's possessing certain moral concepts or making

---

[14] Ibid., 262.

[15] The phrase 'moral judgment' may be ambiguous for the noncognitivists between the linguistic expression of the agent's conative states, or may refer simply to the conative states themselves. So, one might think that an agent judges that "x is good" simply in virtue of having the relevant pro-attitudes towards x, even if she never utters the linguistic expression "x is good."

certain moral judgments. It may be true that I am morally obligated to be generous, but if I am deeply miserly and lack any conception of generosity, or have never made a moral judgment regarding generosity, it seems bizarre to say that I have some motive to be generous merely in virtue of the fact that it's true that I am morally obligated to be generous. So, we need to specify that the necessary connection between one's being obligated to Φ and one's having some motive to Φ must be mediated somehow by one's judgment that one ought to Φ.

So, the existence internalist need not deny the appraiser internalist's view *in toto*.[16] Instead, they will insist that the internalist thesis is to be understood not merely as a thesis about *judgment*. It is a thesis about the nature of, say, moral obligation itself. Moral obligation is such that, if, for example, Jane judges that she ought to tell the truth, and it's true that Jane ought to tell the truth, then Jane also has some motive to tell the truth. We can see the difference between existence internalism and appraiser internalism by considering an example of an agent who lacks any motive to do what they judge that they ought to do. Recall Finn who lacked any motive to keep his engagements. On the appraiser internalist's view, we were forced to conclude in this case that Finn failed to make a genuine moral judgment. This does not follow on the existence internalist's view. If Finn lacks any motive, it might still be the case that he makes a sincere moral judgment that he ought to keep his engagements. If so, then it would then have to be the case that Finn wasn't under an obligation to keep his engagements after all.

Constitutive and Non-Constitutive Existence Internalism

There are two ways of making sense of the above inference on the existence internalist's view. These two ways correspond to two of the different ways in which one might understand the necessary connection between moral judgments and reasons for action. First, one might say that Finn lacks any motive to keep his engagements, even though he sincerely judges that he ought to, because he has failed to actually know or perceive what his moral obligations are. On

[16] Smith, for instance, argues that, on one interpretation of the existence internalist's thesis, they are committed to a view that logically entails appraiser internalism. See Smith 2007, 232.

this view, it is necessarily the case that if Finn did know or perceive his moral obligation, this knowledge or perception would have *caused* him to have some motive. Without the motive, we can infer that, whatever his sincere judgment is about, it is not about his actual obligation. I follow Darwall in calling this version of existence internalism *non-constitutive internalism*:

> Non-Constitutive Internalism: necessarily, if an agent, S, actually epistemically engages with the fact that she ought to Φ, S will, in virtue of that epistemic engagement, be caused to have some motivation to Φ.

A broad version of this view is commonly attributed to Plato, and a narrow view, regarding obligation, can be seen in Richard Price: "when we are conscious that an action…*ought* to be done, it is not conceivable that we can remain *uninfluenced*, or want a *motive* to action" (R 757).[17] What distinguishes non-constitutive internalism from appraiser internalism is that the source of motivation is not provided merely by an ethical judgment.[18] Instead, what provides the motivation is actual engagement with some ethical, or broadly normative, fact. Or, as Darwall characterizes the position, it says, "not that ethical belief or sincere assertion necessarily motivates, but that actual consciousness of or cognitive contact with the ethical does."[19] Darwall does not say much about what this cognitive contact consists in. I think the epistemic state that he has in mind here is something like the epistemic state that intuitionists maintain we are in when we are aware of some *sui generis* moral property. That truths about these properties are self-evident does not entail that they are obvious to anyone who reflects on them. One must be educated to a certain degree, must be thinking clearly, and must be disinterested. Moreover, one can only grasp a *sui generis* property if the property is there to begin with, so there is a success condition built into this epistemic state. The same sorts of

---

[17] Price, Richard. *A Review of the Principal Questions in Morals*. Reprinted in *British Moralists* ed. D.D. Raphael Oxford: Clarendon Press, 1969. v. ii, 757. All references to the Raphael edition will be denominated with an *R* and will be followed by marginal numbers.

[18] This statement presupposes a non-constitutive appraiser internalism. That is, the moral judgment is causally, not constitutively, related to motivation. This seems to be distinct from Stevenson's view above, where the moral judgment is expressive of one's motives.

[19] Darwall 1992, 157.

considerations would seem to be at play in Price's view about obligation. Notice that the success condition is what distinguishes this view from appraiser internalism: our knowledge of our actual obligations provides us with motivation.

The non-constitutive internalist is not trying to give an analysis of ethical terms. Indeed, Price and other intuitionists argue that certain ethical terms do not admit of analysis.[20] However, other internalists do seek to give an account of, say, moral obligation in terms of motivation. These internalists would say that Finn's lacking any motivation to keep his engagements entails that Finn is not actually obligated to keep his engagements because motivation is partly constitutive of moral obligation. Again, I follow Darwall in calling this position *constitutive internalism*:

> Constitutive internalism: necessarily, if an agent, S, is obligated to Φ, then, under certain circumstances, S will have some motivation to Φ in virtue of a constitutive relationship between motivation and obligation.

Different constitutive internalists will understand the nature of this constitutive relationship differently and I will have more to say about this view in later chapters. I am concerned here only to show how the view is different from non-constitutive internalism. The constitutive internalist is attempting to give a partial analysis or definition of moral obligation in terms of motivation, where the non-constitutive internalist is attempting to specify the necessary causal consequences of knowing or perceiving some fact about obligation.

---

[20] However, it's not clear that a term like 'moral obligation' won't admit of analysis even on an intuitionist's view. This is something that is lost in the following argument from Darwall: "Internalism of this sort, however, does not help us understand what normativity itself is. This was no oversight for the intuitionists, since they thought there is nothing to be said about that." (Darwall 1995, 10). This is wrong on two counts. First, intuitionists do think there are things to be said about certain moral terms, like 'good' or 'right.' Just because these terms don't admit of analysis, doesn't mean we can't say anything about them (e.g. that they refer to some *sui generis* moral properties). Secondly, the intuitionist doesn't need to hold that all moral terms are unanalyzable. G.E. Moore in *Principia Ethica* held that, while *good* is indeed indefinable, it is nevertheless the case that *right* admits of analysis. Darwall overlooks this possibility by lumping all ethical terms under the heading of "normativity itself."

Strong and Weak Internalism

Before moving on to discuss the two "internalist" traditions that Darwall locates in early modern British moral philosophy, I note a further way in which internalist positions can be distinguished in order to avoid any misconception about the views under consideration. This consideration will apply to every version of internalism addressed so far. To begin, let's recall the most general statement of motive internalism: necessarily, if an agent, S, is under an obligation to Φ (alternately, if she genuinely judges that she is obligated to Φ), then S has *some* motive to Φ. One might hold a stronger version of this thesis by accepting the following:

> Strong motive internalism: necessarily, if an agent, S, is under an obligation to Φ (alternately, if she genuinely judges that she is obligated to Φ), then S has *a sufficient* motive to Φ.

The idea here is that one's having, or sincerely judging that one has, an obligation brings along with it an all things considered motive to act. If we understand having a sufficient motive as entailing that, *ceteris paribus*, one will so act, then there is no important difference between strong motive internalism and a view that is sometimes called action internalism:

> Action internalism: necessarily, if an agent, S, is under an obligation to Φ (alternately, if she genuinely judges that she is obligated to Φ), then S Φ's.

I mention strong motive internalism and action internalism only to set them to the side. I will be concerned with the weaker version of motive internalism, which holds that there is a necessary connection between moral judgment and an agent's having *some* motivation. If the need arises, I will refer back to the stronger versions of internalism in later chapters.

Empirical Naturalist Internalism

Darwall notes that two different "internalist" traditions emerged in the writings of the British Moralists throughout the 17th and 18th Centuries.[21] He calls the two traditions *empirical*

---

[21] I will henceforth leave off the scare quotes around 'internalism' since I mean to be discussing Darwall's view as he presents it. Again, one of the main problems that I seek to raise in this dissertation is that it is not clear to me in what sense Hobbes, Hutcheson, and Hume all fall within an *internalist* tradition.

naturalist internalism, and *autonomist internalism*. First, we can find a sort of constitutive existence motive internalism in figures that have broadly empiricist epistemologies, and naturalistic metaphysics.[22] Hobbes, Cumberland, Hutcheson, Hume, and Locke all have these sorts of commitments and offer constitutive existence internalist views. Such a view would hold that, since we have no epistemological faculties besides those involved in empirical inquiry, including the capacities for demonstrative, or deductive, reasoning, and there are no moral truths that cannot be explained in fully naturalistic terms (e.g. no super or extra natural facts required), being obligated to act must be understood in terms of motives that are raised (discovered) through the use of theoretical reason (empirically understood). On Hobbes's view, for example, theoretical (or "right") reason is put to the service of ends that are inevitably a part of human being's lives. These ends are understood as the objects of desires, specifically the desire for self-preservation, and the avoidance of pain, and so they are fully naturalistic and amenable to an empirically minded, mechanistic science.

Darwall argues that Hutcheson has a version of internalism that is similar to that of Hobbes's view. However, there is a "difference in emphasis" between the two views.[23] Where Hobbes holds that theoretical reason is merely instrumental in securing ends that are inescapable, Hutcheson holds that a calm and reflective use of theoretical reason can lead creatures with our psychological makeup to desire new ends (see chapters 4 and 7). Both Hobbes and Hutcheson, argues Darwall, agree that theoretical reason has no power of its own to give us new ends. However, Hutcheson seems to think that there are ends, like universal benevolence, that can be discovered by the appropriate use of theoretical reason. According to Darwall, both Hobbes and Hutcheson are committed to a version of constitutive existence motive internalism insofar as

---

[22] Darwall does *not* say that these are versions of constitutive existence internalism. I am using his taxonomy, which he does not apply to his discussion of empirical naturalism, in order to make sense of his claim that this tradition holds that "*obligation consists in motives* raised through the use of theoretical reason" (1995, 14-15; emphasis added).

[23] Ibid.

they maintain the following thesis: "obligation consists in motives raised through the use of theoretical reason."[24] However, they hold distinct views of the capacities of theoretical reason. Nevertheless, they are committed to a view of theoretical reason that does not imbue to that faculty anything outside of what is required for empirical inquiry. Since much of Darwall's statement of these positions is unclear, I will need to return to it in what follows to address its adequacy.

Interestingly, Darwall also locates Hutcheson in a different internalist tradition, *viz*. that version of internalism held by the sentimentalists. Here he groups Hutcheson together with Hume, not Hobbes, and he argues that the sentimentalist account of obligation is not internalist in a strict sense. Understanding the sentimentalist theory in relation to the strictly internalist theories is one of my main aims in chapter 7. This point not withstanding, it is clear that Hutcheson and Hume both share with Hobbes a commitment to a broadly empiricist epistemology and a naturalistic metaphysics. So, they should rightly count as holding a view that is empirical naturalist, even if it is not clear how it could count as a version of empirical naturalist internalism.

Autonomist Internalism

Since the principal figures of this dissertation are all located within the empirical naturalist internalist tradition, I will only give a rough a characterization of the *autonomist* tradition in this introduction. The autonomist tradition can be found in the writings of Cudworth, Locke (sometimes), Shaftesbury, and Butler. As Darwall broadly characterizes the view, it holds that "obligation consists in conclusive motives raised through the exercise of autonomous practical reasoning (that is, the practical reasoning that realizes autonomy)."[25] The main difference between empirical naturalist and autonomist constitutive existence internalist

[24] Ibid.

[25] Ibid.

traditions, then, seems to be that there are certain constraints on the kind of practical rationality that is constitutive of normativity.[26] For the autonomist tradition, having an autonomous deliberative capacity is necessary to raise (create) the ends constitutive of morality. Theoretical reason, for the autonomists, is much more than the means/ends deliberation required by the empirical naturalist tradition.

A Perennial Question: "Why Be Moral?"

Having given very rough sketches of the various versions of internalism that will concern me throughout, I will end this introduction by pointing to a *prima facie* problem that internalist views seem to have. This problem arises when we take seriously the question "why be moral" and so it will be helpful if we first have a rough idea about what that question is thought to be asking.

If we follow Prichard, we might wonder if the question "why be moral" is even intelligible. More precisely, we might think that looking for an answer to this question constitutes a mistake. Prichard argues that a question like, "is there really a reason why I should act in the ways in which I hitherto have thought I ought to act" looks for an answer that it is not possible to give.[27] We might think that we can provide an answer to the question by pointing out that doing what we judge that we ought to do is a sufficient means for bringing about something that is in our self-interest. "The formulation of the question implies a state of unwillingness…and we are brought into a condition of willingness by the answer…for it is just the fact that the keeping of our engagements runs counter to the satisfaction of our desires which produced the question."[28] However, as Prichard points out, this does not in fact answer the original question. We do not gain a reason to do what we judge that we should by appealing to self-interest. We

---

[26] I am not quite sure what Darwall means by saying that obligation *consists in* motives that raised through autonomous practical reasoning.

[27] Prichard 1949, 1.

[28] Ibid., 3.

instead get a reason to do that which satisfies our desires (or otherwise contributes to our self-interest.)[29]

I think there is reason to read Prichard as an internalist.[30] In the course of arguing that the rightness of an action cannot be understood in terms of an action's being done from a particular motive, Prichard says, "To feel that I ought to pay my bills is to be *moved towards* paying them."[31] While establishing internalism is not his main point here, he does seem to be saying that a judgment about an obligation brings with it some motivational state or other, even though he's denying that that motivational state needs to be good. If we read Prichard as an internalist, then we have a straightforward way of making sense of his denying the intelligibility of the question "why be moral?" If by this question one means, "what justifying reason do I have for doing what I judge that I ought to do?" then, on Prichard's view, the question contains within it its own response. That is, if the kind of reason that one is looking for is one which *explains* one's actions (without necessarily *justifying* them), then the judgment that one ought to so act provides just this sort of explanation. The judgment that I ought, for example, to pay my bills brings along with it a motivation to pay them, and so provides me with an explanatory reason for paying them. Asking for a further reason why I should pay my bills is evidence that I did not really understand my moral judgment.[32]

However, one might find this kind of response unsatisfactory. If I am genuinely in doubt as to whether or not I should do what I think that I should do, then it doesn't seem to be of much help to merely be reminded that it is what I think I should do. Something more seems to be

---

[29] Perhaps what Prichard means is that appealing to our desires gives us a motivating reason to do what we should, but that this sort of an appeal does not give us a justifying reason.

[30] However, Darwall notes that Falk reads Prichard as an externalist. See Darwall 1992, 169, n.6.

[31] Prichard 1949, 7.

[32] Alternatively, it is to fail to make a moral judgment at all.

required. What this something more is, or could be, is obviously a matter of some controversy. Still, it seems as though the question requires a more substantive answer.

What is most helpful about Prichard's account is his diagnosis of the original problem, as he sees it: "Any one who, stimulated by education, has come to feel the force of the various obligations in life, at some time or other comes to feel the irksomeness of carrying them out, and to recognize the sacrifice of interest involved."[33] The very intelligibility of the question, then, turns on conceiving of the demands of morality as being something *distinct* from self-interest. That is, at a minimum, it can only be asked if our account of morality presupposes that moral obligations can involve genuine sacrifice. According to Darwall, this sort of genuine sacrifice is absent in traditional natural law theory. Aquinas's version of natural law, for example, holds that "individuals can realize their respective goods only if they function properly in the overall scheme specified by eternal law."[34] Because the eternal law is God's specification of the perfection of all creatures, there is a metaphysical assurance that the ends, the goods, of each individual will, to use Darwall's phrase, "be in harmony."[35] On this sort of a picture, there can be no question that one should do what they truly judge they should do. To do otherwise would not allow them to achieve their own good, and the seeking of good and the avoidance of evil is the fundamental law of nature for Aquinas.

As Darwall tells the story, when Aristotelian metaphysics started to fall out of favor in the 17th Century, so too did the metaphysical guarantee that one's ends would line up squarely with the demands of morality.[36] Without a divinely-backed, teleological guarantee that one's self-interest would be served by following the dictates of morality, it became possible to enquire

---

[33] Prichard 1949, 1.

[34] Darwall 1995, 5.

[35] Ibid.

[36] Ibid. 4-9. See also Darwall 1993, 415-418 for a similar account.

into the rationality of moral action. Darwall's statement of the problem is rather nice, and bears quoting at length:

> A stable harmony of human interests can be achieved only through a natural law (morality) that governs or redirects the pursuit of individual good. But precisely because morality was now thought to involve a form of action guidance distinct from prudence, it became thinkable that individuals might do better by not being moral, whether occasionally or throughout life. Thus the modern problem: how does morality obligate? (Darwall 1992, 416)

This is a question famously dealt with by Hobbes in his response to the fool. It is a question that Plato posed at the beginning of the *Republic*, and only after Aristotle's influence began to wane did it start to seem again like an important question to ask.

Yet, we saw in Prichard that there seems to be a tension between holding a version of internalism and taking seriously what Darwall calls the modern moral problem. Brink puts the problem into the form of a dilemma regarding the possibility of the amoralist, or the intelligibility of the amoralist's demands. If we accept a version of internalism about motives, then the amoralist seems to be conceptually incoherent. The amoralist, as Brink defines him, "is someone who recognizes the existence of moral considerations and remains unmoved."[37] Also, if we accept a version of internalism about justifying reasons, then an amoralist who asks the question "why should I be moral" has created a purported challenge that is in fact based on a conceptual confusion.[38] Brink does not think that the amoralist is conceptually incoherent, and does not take the amoralist's challenge to be obviously rooted in conceptual error. So, he rejects both internalism about motives and internalism about reasons.[39]

In the context of the moral problem of the 17th Century rejection of teleological metaphysics, this form of argument against the internalist thesis should give us pause. According to Darwall, one of the fundamental questions that moral philosophers took themselves to be

---

[37] Brink 1989, 46.

[38] Ibid., 59.

[39] For a similar argument against internalism, see Tenenbaum 2000, 108-130.

addressing was, to quote Shaftesbury, "what obligation is there to virtue; or what reason to embrace it" (R 205). Nonetheless, Darwall attributes to these same figures various internalist theses. If we take Brink's argument seriously, as I think we should, then there is at least a *prima facie* inconsistency between one's holding a version of internalism and thinking that the question "why be moral?" merits a substantive response, or is, at a minimum, an intelligible question. I think that attempting to locate internalist commitments in Hobbes, Hutcheson, and Hume is a worthwhile project precisely because of this tension: if they are engaged in the modern moral problem, as I assume for the time they are, then what would they have to say in response to this sort of *prima facie* tension?

CHAPTER 1

HUMAN NATURE, MORALITY, AND THE PROBLEM OF THE FOOL

Introduction

There are a few reasons why I am beginning this dissertation with a discussion of Hobbes's moral theory. Since one of my principal aims is to understand the relationship between internalism and the question "why be moral" it will be helpful to begin with a figure who, I argue, accepts internalism, and who also has an explicit discussion of the question. The discussion of this question comes from the fool's objection and Hobbes's response to it in Chapter xv of *Leviathan*. So, understanding the fool's objection and Hobbes's response to it will be a principal aim of this chapter. Also, since Hobbes, Hutcheson, and Hume are all part of the empirical naturalist tradition, it is important to get an understanding of the historically prior theory in order to trace its development through its later, explicitly sentimentalist, developments. Finally, Hobbes is an important transitional figure between the traditional natural law theorists, and the early modern British Moralists. Much of Hobbes's technical terminology is taken from the natural law tradition, most obviously his use of the expression 'laws of nature'. However, Hobbes provides a fully naturalistic understanding of these terms so that they are amenable to a mechanistic science that does away with teleology. So, understanding Hobbes is crucial for understanding the sort of change in moral philosophy that Darwall points to as leading to the "modern moral problem."

In this chapter, I first present Hobbes's conception of the right of nature as it features into his account of human beings in the natural condition. I will also describe the laws of nature and the mechanism by which we are, supposedly, capable of escaping the state of nature. I will then present the fool's objection and Hobbes's response to it. I show a number of different interpretive options that have emerged from the secondary literature on this passage. I do not here argue in favor of any of them. Instead, I provide them because I think they make clear the various interpretive issues that need to be addressed in order to understand Hobbes and the fool. I

conclude this chapter with a discussion of Hobbes's understanding of moral philosophy and ethics in relation to his division of the sciences.

Rights and the First Two Laws of Nature

Hobbes raises the fool's objection at the end of his discussion of the third law of nature prescribing to people to keep the covenants they make. Since the third law of nature, like all other laws of nature, is derived from the first and fundamental law, roughly to seek peace, and since the discussion depends on Hobbes's particular conception of notions like "right" and "law," I will first describe some of the relevant context for understanding the fool.

For Hobbes, all agents have an inviolable right of nature. I clarify the sense in which this right is inviolable below. To begin, he defines the right of nature as follows:

> the liberty each man hath to use his own power, as he will himself, for the preservation of his own nature, that is to say, of his own life, and consequently of doing anything which, in his own judgment and reason, he shall conceive to be the aptest means thereunto…By Liberty is understood, according to the proper signification of the word, the absence of external impediments, which impediments may oft take away part of a man's power to do what he would, but cannot hinder him from using the power left him, according as his judgment and reason shall dictate to him (L xiv.1-2, R 55).[1]

Famously, Hobbes argues that the natural condition, a condition where there is no common power set above individuals living in close proximity to each other, is a condition of war of all against all. The details regarding the causes of this war need not immediately concern us here. What is important about the natural condition here is the way in which Hobbes describes the behavior of those individuals who unhappily find themselves in such a condition. He says that it is reasonable for these individuals to engage in preemptive attacks for the sake of greater

---

[1] Thomas Hobbes, *Leviathan with selected variants from the Latin edition of 1688*, ed. Edwin Curley (Indianapolis: Hackett Publishing Company, 1994). Text references to *Leviathan* are preceded by *L* and are followed by chapter and paragraph of the Curley edition. Where possible I include reference to D.D. Raphael's *British Moralists 1650-1800* (Indianapolis: Hackett Publishing Company, 1991). These references are preceded by 'R' and are followed by the marginal numbers of the Raphael. We might, following Bernard Gert, understand "liberty" here to mean not only the absence of external impediments, but a freedom from a particular sort of obligation. Individuals in the state of nature are free from the sort of obligation that follows from the making of contracts. See Gert 1991, 3-32.

security. Such preemption is "no more than his own conservation requireth," and so this "dominion over men…ought to be allowed him" (L xiii.4, R 48). This *ought* cannot be understood as a moral ought since "the notions of right and wrong, justice and injustice, have there [in the natural condition] no place" (L xiii.13, R 53). The idea here is that the dominion of one person over another in the natural condition cannot be evaluated as being either morally right or morally wrong. However, it can be evaluated as being rational wherever such dominion is, or is judged to be, required for an agent's self-preservation. So, we can understand the claim that individuals ought to be allowed dominion over each other as following from the claim that individuals have a right to such dominion where it is, or is judged to be, important for that individual's self-preservation. According to this conception of rights, then, it does not follow that an individual's right to *x* entails an obligation on others not to interfere with her obtaining or participating in *x*.[2] One has a right to dominion in the natural condition, but it's not the case that others are obligated to refrain from preventing such dominion over themselves: the claim that such dominion ought to be allowed cannot be understood as the claim that others ought to allow themselves to be dominated.[3]

So, in the natural condition, individuals have a right to dominion over each other (where this dominion could, and likely will, involve destruction of the other). More generally, Hobbes argues that there is nothing that *in principle* might not figure into an individual's judgments about what is required for her self-preservation.[4] Besides possessing a will, the only requirement for an individual to have a right to *x* is that she judge that *x* is the "aptest means" to achieve her

---

[2] As Warrender notes, Hobbes's conception of a right does not entail corresponding duties or obligations on others. So, for Hobbes, it is not the case that if S has a right to *x*, then Q has a duty not to interfere in S's obtaining or participating in *x*. An individual in the natural condition has a right to whatever she judges to be conducive to her self-preservation. If two individuals make the same judgment with respect to the same object, they will both have a right to it, but neither will have an obligation to not interfere with the other's obtaining it. See Warrender 1957, 18-21.

[3] This is because, as we'll see below, Hobbes argues that individuals can never lay down their right to self-defense.

[4] For a helpful discussion of this point, see Kavka 1986, 300-2.

own self-preservation.[5] Hobbes doesn't seem to think that an individual's judgments about aptest means need to be accurate in order to ground the right of nature. It needn't be the case that if S mistakenly judges that *x* is the aptest means to her self-preservation, then S doesn't truly have a right to *x*. So, when he says, "it followeth that in such a condition every man has a right *to everything*, even to one another's body," he means to say that that to which an individual has a right is determined by that individual's own judgment that it is important for her self-preservation, regardless of her epistemic justification for that judgment (L xiv.4, R 57; emphasis added). Of course, given the severity of the natural condition, it behooves us be concerned with forming accurate judgments about how we can go about staying alive.

For individuals who are primarily concerned with their self-preservation, such a condition is obviously not desirable. It is not clear whether Hobbes thinks that human beings are *only* capable of pursuing, ultimately, their own self-preservation. Psychological egoism, as I understand it here, is the view that, as a matter of psychological fact, individuals are only capable of desiring, as an end, their own self-interest.[6] I deny that 'desire' is to be understood as whatever it is that motivates one to act, and that 'self-interest' is to be understood as the satisfaction of desire. On that understanding, the view would be trivial, asserting that agents always act from whatever motivates them to act.[7] I follow Gregory Kavka thinking the most interesting version of the thesis is a nonmaximizing version of egoism: the view does not hold that agents always pursue what is, on balance, most in her long-term interest, but is only capable

---

[5] By "aptest means" I think we must interpret Hobbes as meaning "best means currently available." The evaluation of those means that are the best should include considerations of risk. On this interpretation, there may be courses of action that lead directly to a large quantity of goods for me, but because these courses of action are incredibly risky, I have a right to pursue other courses of action that lead either indirectly to the same quantity of goods, or directly to a lesser quantity of goods.

[6] It is clear that both Hutcheson and Hume took Hobbes to maintain this view. For Hutcheson's reading of Hobbes as a psychological egoist, see *Inquiry* I.iii, R 310; for Hume's reading see EPM appendix 2, SBN 297.

[7] Kavka 1986, 35.

of pursuing ultimately something that is in her interest.[8] Moreover, there are number of ends that could count as being in an agent's interests, so I do not think that psychological egoism is essentially hedonistic.[9] Hobbes claims that "of the voluntary acts of every man the object is some *good to himself*" (L xiv.8, R 60). However, the phrase "good to himself" is ambiguous. It could mean, "whatever it is that he happens to desire." On this interpretation, Hobbes need not be committed to psychological egoism, because it leaves open the possibility that an individual desires, as an end, the good of another. However, if "good to himself" is understood as "a personal benefit" then the psychological egoist reading clearly follows.[10]

There is not much compelling textual evidence to support the claim that Hobbes accepted psychological egoism.[11] However, the case is not entirely clear. For my purposes here, however, all that needs to be seen is that Hobbes does maintain that individuals are concerned, ultimately, with their own self-preservation:

> For every man is desirous of what is good for him, and shuns what is evil, but chiefly the chiefest of natural evils, which is death; and this he doth by a certain impulsion of nature, no less than that whereby a stone moves downward. It is therefore neither absurd nor reprehensible, neither against the dictates of true reason, for a man to use all his endeavours to preserve and defend his body and the members thereof from death and sorrows. (DCv I.7)[12]

---

[8] Ibid. 38-9.

[9] For further discussion see Kavka 1986, 35-44. There are some ends that are incompatible with psychological egoism, among them the well-being of others for its own sake.

[10] Ibid. This discussion is admittedly rather brief. Gert notes that psychological egoism "though often explicitly attacked, seems to be one of those views which no philosopher explicitly defends" (1967, 505). He takes it that this is part of the reason why it is difficult to explain what psychological egoism amounts to. The way that I have defined the view, however, should work for my purposes here since it is neither a trivial claim, nor is it a thesis that is either obviously held or rejected by Hobbes.

[11] For the arguments against reading Hobbes as an egoist, see Gert 1967, 1991, and 2001; Rhodes 1992 and 2002; and Kavka 1986, chapter 2. Of these authors, Kavka is the only one that maintains that Hobbes did endorse some non-psychological version of egoism. Rhodes explicitly denies this point, and Gert is unclear on the issue, although he does reject the psychological egoist reading.

[12] References to both *De Cive* and *De Homine* come from Gert's *Man and Citizen* (Indianapolis: Hackett Publishing Company, 1991). References to *De Cive* will be abbreviated DCv and will be followed by chapter and paragraph number; references to *De Homine* will be abbreviated DH and will be followed by chapter and paragraph number.

This passage does not show that Hobbes is committed to psychological egoism since, again, the phrase "good for him" is ambiguous. Hobbes thinks that something's being good for someone is just a matter of her desiring it (L vi.7, R 25). To say that we all desire what is good for us is, for Hobbes, simply trivial. His claim that we desire to avoid death by an impulsion of nature does not entail that we always pursue our own self-preservation any more than the his claim that stones move downward by an impulsion of nature entails that stones always move downward. Hobbes's main point in this passage is that just as the motion of stones is a natural force governed by the laws of nature, so too is the pursuit of one's own self-preservation a natural force consistent with the laws of nature. The end of self-preservation is entirely consistent with the dictates of reason, which, as we'll see below, are the laws of nature. This lends support to the claim that Hobbes accepts a version of rational, not psychological, egoism.[13] Actions that are done in the service of this end, then, cannot, except in the rarest cases, be thought to be irrational.

There are other places where Hobbes makes claims that seem to lend support to reading him as a psychological egoist.

> The greatest of goods for each is his own preservation. For nature is so arranged that all desire good for themselves. Insofar as it is within their capacities, it is necessary to desire life, health, and further, insofar as it can be done, security of future time. On the other hand, though death is the greatest of all evils (especially when accompanied by torture), the pains of life can be so great that, unless their quick end is foreseen, they may lead men to number death among the goods. (DH xi.6)

In this passage we see, first, that Hobbes takes self-preservation to be the greatest of goods. Then it seems as though Hobbes goes on to assert psychological egoism insofar as he claims that all men desire "good for themselves." However, this impression is misleading since he goes on to say only that what is desired are life, health, and security. Without life, desire is impossible.[14] Poor health and insecurity put limits on our power to obtain what we desire. Nothing in this

---

[13] The view these comments suggest is that there is nothing irrational about the pursuit of one's self-preservation. More generally, this version of rational egoism would hold that it is never irrational to pursue one's self-interest.

[14] See L xi.1

passage, then, suggests psychological egoism since it puts no restriction on what an individual may desire as an end. Finally, this passage points to the sort of rare, limiting case in which, Hobbes seems to suggest, the aim of self-preservation is not rational. It is only when one's life has become so painful and wretched that one can think of death as a good, and self-preservation as an evil. Note, even in this case, Hobbes will still maintain that the individual aims at some good to herself, *viz.* the alleviation of suffering.

Given the (typical) importance of self-preservation, and the great uncertainty that this end can be realized in the natural condition, Hobbes thinks that individuals will reason that there are better arrangements and will also discover through reason in what those better arrangements consist. This explains why, in the introduction to his discussion of the laws of nature in *Leviathan*, Hobbes says that "reason suggesteth convenient articles of peace, upon which men may be drawn to agreement" (L xiii.14, R 54). Also, Hobbes variously refers to these articles of peace, that is the laws of nature, as "general rules" of reason, and as "dictates of reason" (L xiv.4, R 56; L xv.41, R 77). Another rather clear example of this is in Hobbes's definition of the law of nature in *De Cive*: "the *law of nature*, that I may define it, is the dictate of right reason, conversant about those things which are either to be done or omitted for the constant preservation of life and members, as much as in us lies" (DCv ii.1).

So, according to a common reading of Hobbes, by the use of the very same kinds of judgments and reasons that lead us to a condition of war (namely those that take as their principal aim an individual's self-preservation), Hobbes argues that we will be able to escape from that condition. The same kinds of reasons that make it rational for us to attack preemptively, seek dominion over others, and, in general, use what we judge to be the aptest means to our preservation in the state of nature, so too should we follow the rational dictates encoded in the laws of nature.[15] The end is the same in both cases, and presumably reason

---

[15] Kavka argues that this common reading assumes that rationality is a matter of case-by-case expected utility maximization. Since rationality in the commonwealth is not a matter of case-by-case expected utility maximization, this claim about "for the same reasons" is false. I will later argue that something like Kavka's reading is correct. Kavka 1995, 23.

fulfills the same function in both, namely dictating those means that are the most apt for achieving the end.[16] So, what accounts for the difference between following reason to a condition of war and following reason to a condition of peace? For Hobbes, the difference could only be external to the agent herself. It is only under certain external conditions that it becomes rational for one to pursue courses of action that lead to peace as opposed to those that lead to war. Primarily, Hobbes appeals to certain facts about other agents to account for the relevant external difference. That is, Hobbes argues that it is only when others are willing to constrain their behavior in particular ways that makes it rational for us to pursue cooperation, and irrational to pursue war. While I do need to believe that others are so willing, the primary shift is external and occurs in the willingness of others to cooperate. Similarly, when others are not willing to constrain their behavior in particular ways, then it is rational to pursue war and irrational to pursue peace.

This external difference is built into the formulation of the first law of nature, which states "that every man ought to endeavour peace, as far as he has hope of obtaining it, and when he cannot obtain it, that he may seek and use all the helps and advantages of war" (L xiv.4, R 57). Again, 'ought' here cannot refer to what agents ought morally to do, but it is an 'ought' of practical rationality. It is rational for agents to seek peace only when they have hope of obtaining it. "Hope," as Hobbes defines it in *Leviathan*, is an "appetite with an opinion of attaining" (L vi.14, R 28).[17] The most straightforward way of understanding the hope for peace, then, is to understand it as both a desire for peace and an opinion that such peace is possible. Hobbes seems to be saying that it is rational for agents to pursue peace in the state of nature only under certain

---

[16] As we'll see below, there is some controversy about whether or not Hobbes thought that reason was only capable of revealing means to ends, or if reason was able to supply ends and goals of its own.

[17] It is odd that Hobbes defines hope in terms of appetite when just a few paragraphs earlier he had distinguished appetite from desire "the latter being the general name, and the other oftentimes restrained to signify desire of food" (L vi.2, R 24). Hobbes does think we can hope for things besides food, namely, in this case, peace, and so it seems desire would have been the more appropriate term. In what follows, I will ignore this technical distinction and will treat 'appetite' and 'desire' as synonyms.

conditions, namely when they both desire peace and when they have an opinion that peace is possible.

Hobbes seems to suggest that this opinion that peace is possible be supported by epistemic reasons. An agent has an epistemic reason to believe that *p* (or judge that *p* in Hobbes's parlance) if she has considerations available (in some sense) to her that support the truth of *p*.[18] Now, it seems clear that an agent has a practical reason to pursue peace only if, in addition to desiring peace, she has an epistemic reason for judging that that peace is possible.[19] What it is rational for an agent to *do* is, in part, a function of what she has reason to *judge* is likely to promote her self-interest. It would be practically irrational for an individual to seek peace without having any epistemic reasons for thinking that peace is possible. The sense of 'possible' here does seem to be something more than simply not-impossible, that is, something more than there being no absurdity in the name "peaceful society of men."[20] The first law of nature dictates that agents seek peace when they have hope of obtaining it, not merely when they desire peace and think there's nothing absurd about such a desire. The opinion that peace is possible is, I think, the opinion that peace is to some degree likely. This, however, isn't saying much since, on one interpretation of *likely*, it's just saying the same thing as *not inconsistent* (since any claim that is not necessarily false is in some sense likely). The agent, it seems, must have epistemic reasons to think that peace is a state of affairs in which she will likely find herself. But how likely such a state of affairs must be, and what evidence she must have for thinking that it is likely, are questions that Hobbes unfortunately never addresses.

---

[18] I roughly follow Diane Jeske's formulation of epistemic reasons for belief. See Jeske 2008, 5.

[19] Hobbes uses 'opinion' here and not 'belief,' since he takes opinion to be the more fundamental term. 'Belief' is defined in terms of opinion, and, for Hobbes, also essentially involves testimony. See L vii.5. I will below use 'judgment' in place of 'belief' in ordinary philosophical locutions such as 'justified belief'.

[20] See L iv.20-1 for Hobbes on insignificant speech.

If the first law of nature commands agents to seek peace, when they have hope of attaining it, then the second law of nature specifies the primary means by which such peace is to be achieved. The second law of nature says "*that a man be willing, when others are so too, as far-forth as for peace and defence of himself he shall think it necessary, to lay down this right to all things, and be contented with so much liberty against other men, as he would allow other men against himself*" (L xiv.5, R 58).[21] The means to peace, Hobbes claims, lie in the notion of laying down one's right to all things, and, to get an understanding of this, Hobbes presents us with the technical machinery that structures the move out from the natural condition, viz. contracts and covenants. Hobbes's discussion of contracts and covenants is important for two reasons here. First, it is in connection with the notion of laying down a right that Hobbes defines the notions of being obliged or bound. Secondly, it is in the same connection that Hobbes first says what it is for agents to act unjustly. So, it will be useful to spend some time considering the nature of contracts and covenants.

As we saw, the second law of nature commands individuals to lay down their right to all things. The second law of nature still admits of exceptions, much as the first law did, in cases where certain features of the world external to the agent obtain. Namely, when other agents are unwilling to lay down their right to all things, Hobbes says, "then there is no reason for anyone to divest himself of his [right to all things]; for that were to expose himself to prey (which no man is bound to), rather than to dispose himself to peace" (L xiv.5, R 58). The mere fact that others are willing to lay down their right to all things is not by itself sufficient for an agent to have a reason to lay down his or her own right. The agent must also have some indication or mark that this is in fact the will of others. Without such an external mark, the mere fact that

---

[21] Hobbes says, "it hath been showed…that the law of nature commands every man, *as a thing necessary*, *to obtain peace*, to convey certain rights from each to other" (DCv iii.1; emphasis added). The laying down of rights is, I take it, necessary only relative to certain facts about individuals in the natural condition. That is, for individuals who are rationally interested in their own self-preservation, who are roughly naturally equal in strength and intelligence, and who live in conditions of scarcity, the laying down of rights is the only means by which they can establish lasting peace.

others are willing to divest themselves of their right provides no agent with a reason to forfeit his or her own right. Hobbes says that the signs by which others indicate their will to contract are "either *express* of *by inference*. Express are words spoken with understanding of what they signify," and signs by inference include silence in certain situations, actions in certain contexts, inaction in other contexts, and, in general "whatsoever sufficiently argues the will of the contractor" (L xiv.13-14, R 62). There are a number of difficulties in specifying how these external signs could ever be taken as signs of another's will when we all know full well that the signs may exist even in an uncooperative will: "the bond of words are too weak to bridle men's ambition, avarice, anger, and other passions" (L xiv.18, R 63). Since figuring out in what this mark could consist would take us too far afield, I will here stipulate that such a mark is present.

When the will of others to lay down their right to all things is known, or when an agent reasonably judges that this is the will of others, then the second law of nature commands agents to lay down their right to all things. This means that each agent is to "divest himself of the liberty of hindering another of the benefit of his own right to the same" (L xiv.6, R 59). Again, Hobbes understands liberty to be "the absence of external impediments" (L xiv.2, R 55). So, Hobbes holds that agents in full possession of the right of nature are to each other external impediments to their achieving their ends. Whether this is so of necessity need not concern us for the moment.[22] The important point here is that to lay down a right is simply to remove oneself as an impediment to someone else's achieving his or her ends. But this account is very hard to understand. Hobbes defines the right of nature in terms of liberty and liberty in terms of the absences of impediments. However, Hobbes also wants to say that every individual stands to one another as an external impediment in the natural conditions. That we are surrounded by external impediments in the state of nature suggests that we do not have liberty in the state of nature which in turn suggests that we do not have a right of nature.[23] Moreover, the problem cannot be

---

[22] I think there is a sense in which we are *necessarily* impediments to each other. I think that such a condition is necessary relative to certain facts about human beings and their external environment.

[23] I thank Diane Jeske for bringing this rather serious problem to my attention.

solved by turning to Hobbes's account of the nature of the external impediments in the *Leviathan* where he says, as we've seen, that they "may oft take away part of a man's power…but cannot hinder him from using the power left him" (L xiv.2, R 55). No matter how Hobbes understands the external impediments, the difficulty arises because he simultaneously asserts that such impediments are present in the natural condition and that liberty is understood as an absence of such impediments. I not sure whether Hobbes can deal with this problem.

In any case, it is clear that Hobbes thinks that agents in full possession of the right of nature are to each other external impediments to their achieving their ends. So, the question arises, in what sense are individuals capable of removing themselves as external impediments to each other's ends? The literal interpretation would seem to have it that individuals must physically restrain themselves: I would obviously present no external impediment to your pursuit of your ends were I to lock myself in a crate. If there are no other alternative interpretations of what it means for an individual to divest herself of a liberty, then we would have to understand Hobbes as making this rather unfortunate claim.[24] However, there are more plausible alternative interpretations of this point. Gert, for example, maintains that liberty, in the sense required for the statement of the right of nature, is a freedom from contractual obligations. Since the natural condition is a state where, it seems, there are no contractual arrangements, it follows that in this state there is freedom from contractual obligation. In this condition, individuals are free to pursue whatever aligns with their own reason and judgment. "Thus," says Gert, "natural right is best defined as freedom from contractual obligation."[25] If Gert's interpretation is correct, then we can understand the claim that an individual divests herself of her liberty as the claim that she takes upon herself a contractual obligation.[26]

---

[24] As Rosamond Rhodes rightly points out, "by renouncing a liberty no one either gains or loses any physical liberties because, physically, no one can do any more or less after an act of laying aside a liberty" (2002, 52).

[25] Gert 1991, 19.

[26] If Gert's interpretation is to be satisfactory, he needs to explain how it can be reconciled with Hobbes's explicit definition of 'liberty' as "absence of external impediments." Being free from

Individuals can divest themselves of liberty by either renouncing or transferring their rights, but the technical distinction between the two is not particularly relevant here,[27] and this is because in either case, Hobbes says, when the right has been laid aside,

> then is he [the agent who laid down his right] said to be Obliged or Bound not to hinder those to whom such right is granted or abandoned from the benefit of it; and [it is said] that he *ought*, and it is his Duty, not to make void that voluntary act of his own, and that such hindrance is Injustice, and Injury, as being *sine jure*, the right being before renounced or transferred. (L xiv.7, R 59)

By laying down a right one is thereby obliged not to interfere in the course of others pursuing their ends, since to lay down a right just is to be obliged or bound. Moreover, to interfere would be unjust. Given that, at this point in the argument, Hobbes has made no mention of a sovereign, it seems as though he maintains that there are obligations in the natural condition, and that, after rights have been laid aside, there are also some actions considered unjust. This is one of the interpretive points at issue in the debate over how to understand the fool's objection, and so we will return to it below.

When rights have been mutually transferred, meaning that two (or more) parties lay down their right to something with the intention that the other(s) should receive the benefit thereby, then the parties have entered into a contract. If the contract is such that one party is to perform before the other party, then the contract is called a covenant (L xiv.11, R 61). Of particular interest to us here are the conditions under which covenants may be rendered void, and hence are not cases where an agent has successfully laid down her right. To begin, it is important to note that Hobbes sets restrictions on the rights that can be subject to being set aside. It is not the case

---

contractual obligations is neither necessary nor sufficient for being free from external impediments. However, if we understand 'external impediments' here to include sanctions for contract violations, then it will turn out to be the case that assuming a contractual obligation brings with it a certain sort of external impediment. Rhodes makes a similar interpretive move and claims that the creation of obligation comes from the voluntary self-divesting of liberties and the creation of *internal* constraints.

[27] A right is renounced when it is given up but the agent is indifferent to who is to receive the benefit. A right is transferred when the benefit of the forfeiture is intended for a particular person or persons. See L xiv.7, R 59.

that people can voluntarily transfer any right they choose. The act of laying down a right is a voluntary act and will therefore always be constrained by Hobbes's conception of the relation between goods and the will: "the proper object of every man's will is some good to himself" (L xxv.2, R 85). If the abandonment or transference of a right cannot result in some good for the agent him or herself, then such an abandonment or transference cannot be voluntarily done.[28] He says:

> And therefore there be some rights which no man can be understood by any words or other signs to have abandoned or transferred. As, first, a man cannot lay down his right of resisting them that assault him by force, to take away his life, because he cannot be understood to aim thereby at any good to himself. [Second], the same may be said of wounds, and chains, and imprisonment…[Third] and lastly, the motive and end for which this renouncing and transferring of right is introduced, is nothing else but the security of a man's person, in his life and in the means of so preserving life as not to be weary of it. And therefore if a man by words or other signs seem to despoil himself of the end for which those signs were intended, he is not to be understood as if he meant it, or that it was his will. (L xiv.8, R 60)[29]

This is the sense in which Hobbes thinks that agents have some inviolable rights; they, unlike other rights, cannot be laid down in contracts or covenants. The reason for this seems to be that the only reason agents have for laying down rights via contracts is for their own good. Hobbes supposes, plausibly perhaps, that no good can come to the agent by allowing herself to be attacked, thrown in prison, or be held to contracts she falsely and ignorantly thought to be of no serious harm. I think that Hobbes is suggesting here that even if someone were to try to lay down her right to self-defense, say, it would almost never be reasonable to count on someone following through. Only in the most extraordinary and rare cases would an individual not resist bondage, torture, and death. Hobbes claims that covenants "wherein neither of the parties

---

[28] The 'cannot' here needs elucidation. Does Hobbes think it's contradictory for a person to lay down her right to, say, self-defense? Or does he hold that we are just psychologically compelled to defend ourselves in apparently life-threatening situations? Or is giving up the right to self-defense just irrational?

[29] I think that Hobbes is suggesting here that even if someone were to try to lay down her right to self-defense, it would almost never be reasonable to count on someone following through. Only in the most extraordinary and rare cases would an individual resist bondage, torture, and death.

perform presently, but trust one another, in the condition of mere nature…upon any reasonable suspicion it is void" (L xiv.18, R 64). Since it is almost always reasonable to expect others will act on their own judgments concerning their self-interest, it will almost never be reasonable to depend on someone acting contrary to these judgments.[30] Since rational agents know that the pursuit of these ends is not to be hoped for, they also know that an individual attempting to enter into one of these agreements should not be taken at his or her word. These purported agreements are always void because no rational party would enter into them with even the slightest understanding of human nature.

The claim that reasonable suspicions renders void any covenant is very important for Hobbes. Just as in the natural condition there is nothing in principle that an individual might not determine to be conducive to her self-preservation, so too there doesn't seem to be any in principle constraint on what counts as reasonable suspicion. There is, however, one caveat, "The cause of fear which maketh such a covenant invalid must be always something arising after the covenant made" (L xiv.20, R 64). The idea here is that if there was nothing preventing someone from entering into a covenant at the time it was made, then, if the agent discovers nothing new regarding her contractual partner or their agreement, it shouldn't all of a sudden become irrational to perform. Nevertheless, there is in principle no new information that an agent cannot reasonably take to indicate an unwillingness of the other party to perform:

> For he that performeth first has no assurance the other will perform after, because the bonds of words are too weak to bridle men's ambition, avarice, anger, and other passions, without the fear of some coercive power, which in the condition of mere nature, where all men are equal and *judges of the justness of their own fears*, cannot possibly be supposed. (L xiv.18, R 64)

This suggests that, in the natural condition, there are no valid covenants. However, the concern here applies directly to covenants of mutual trust, where neither party performs presently. It

---

[30] The individual who falsely believes that no serious harm will come to her through keeping a particular agreement presents some difficulty on this interpretation. I think Hobbes wants to say that, given rough natural equality, so long as it's possible that she discovers her mistaken judgment, it's not rational to rely on her purported agreement.

appears as though it cannot apply to covenants generally since, where one party performs immediately there can be no reasonable suspicion on the part of the other party that the first might not perform. The ubiquity of reasonable suspicion seems only to invalidate covenants of mutual trust, and so doesn't seem to show that there are no genuine covenants in the natural condition.

In fact, Hobbes himself suggests that there are some valid covenants in the state of nature. These are covenants entered into out of fear in the natural condition. Hobbes argues that such covenants are valid and obligatory, and his reasoning seems to be similar to the above argument. Hobbes discusses a case where an agent covenants to pay a ransom to her enemy in exchange for her life (L xiv.27, R 65). He seems to think that by allowing her to live, the enemy has fulfilled his part of the covenant immediately and so there can be no reasonable suspicion on the part of the captive that her enemy will fail to perform. So, the condition that invalidates covenants does not apply to this case and it looks as though her agreement to pay the ransom is obligatory. Hobbes goes on to say that, "even in commonwealths, if I be forced to redeem myself from a thief by promising him money, I am bound to pay it, till the civil law discharge me" (ibid.). It seems that what makes these agreements binding is that, at least temporarily, some other party has sufficient power to compel performance. However, there is no assurance that this power will be lasting in either the natural condition or in the commonwealth. The thief in the commonwealth is less powerful than the sovereign and its enforcement mechanisms, and one's obligation to pay the thief is discharged when the more powerful entity no longer compels my performance. Rough natural equality and the passage of time provide similar consideration in the case of covenants of fear in the natural condition.

To see this, suppose that the agent covenanted to pay a ransom to her enemy in exchange for *continued* life (i.e. beyond the end of ransom transaction). Now, when the agent brings back the ransom to her enemy, she may have no reason to suppose he won't kill her as soon as he has

the ransom.[31] In this case, her enemy has not yet performed his end of the bargain because their bargain extends beyond the time where end of the ransom transaction. Hobbes cannot appeal to the idea that whatever was rationally covenanted cannot rationally be forgone. This is because the covenant was entered into by fear, or duress.[32] The agent may think, "I will promise to pay the ransom so that I will not be killed immediately." Moreover, she may think this while simultaneously thinking, "It's very likely my enemy will kill me once I pay the ransom." Nevertheless, it seems reasonable for the agent to make the covenant, but not to keep it if she can get away with not doing so.[33] The sense in which the covenant with her enemy is valid seems just to be that her enemy has power enough to hold her to her word. If she can, however, find some way of evading that power, then there doesn't seem to be reason for her to keep her agreement especially when doing so would further jeopardize her security. Given rough natural equality, and the time difference between her captors initial performance and her "required" performance, it seems likely that she will find such a means to evade her captors power. At most, then, what Hobbes's discussion of valid natural condition covenants shows us is that there will be very few such covenants in the natural condition. However, perhaps it is mistaken to think that if we can get away with violating an agreement, and that doing so is to our advantage, Hobbes would say that violating the agreement is rational. This is one of the issues that arises in Hobbes's dispute with the fool.

---

[31] I stipulate here that the ransom transaction is not a theft for two reasons. First, it could not be a theft since there is no personal property in the natural condition. Second, and more importantly, Hobbes says that *covenants* are valid and obligatory in the natural condition when they are entered into by fear. So, we need to stipulate that the case does not involve the simultaneous and immediate performance by both parties.

[32] I think that one of the main problems underlying those attempts to say that obligations are grounded in the use of certain words is that they do not pay enough attention to the role that passions play in our use of certain expressions. See Parry 1967, 249; and Peacock 2010, 441.

[33] I think Peacock is right to point out that there is an important difference between reasons and motives for entering into covenants and reasons and motives for keeping covenants once entered into. See Peacock 2010, 434-7.

The Fool's Objection and Hobbes's Response

So, the first law of nature dictates to individuals to seek peace, where there is hope of attaining it, and the second law of nature instructs individuals on the way in which peace is to be obtained, namely through the laying down of rights via contracts and covenants. However, we've seen that our ability to use covenants in the natural condition is rather limited. The making of covenants is a useless method for establishing peace if no one bothers keeping the covenants they've made. So, Hobbes concludes that there is a further law of nature, the third law, which requires "*that men perform their covenants made*" (L xv.1, R 67). Hobbes here elaborates on a point he made about injustice in his discussion of the second law of nature: "And in this [third] law of nature consisteth the fountain and original of justice…when covenant is made, then to break it is *unjust*; and the definition of injustice is no other than *the not performance of covenant*. And whatsoever is not unjust, is *just*" (L xv.2, R 67).[34] As we have seen, Hobbes argues that there can be no justice or injustice in the natural condition, and we can now see why. Injustice is the non-performance of a covenant, and in the natural condition covenants are rendered invalid so long as there is reasonable fear of non-performance. So, what's needed in order to secure the rationality of keeping covenants is a common power that is capable of compelling performance. But, there can be no such power until the establishment of a commonwealth, and if commonwealths can only be established by valid covenants of mutual trust, then it seems as though the rational means we have for escaping the natural condition are self-defeating. But this problem is not of immediate concern for us.

I will here take for granted that the creation of a commonwealth is both possible and rational, because the fool's objection seems to presuppose both.[35] Since the passage is

---

[34] I do not think that Hobbes means to say that all actions that are not violations of covenants are just. This implies that all actions in the state of nature are just merely because there are (almost) no covenants to be broken in the state of nature. I think that Hobbes intends to define justice as the keeping of valid covenants.

[35] Since the objection questions the rationality of *injustice*, and injustice depends on the existence of a common power, the fool seems to presuppose the existence of a commonwealth. Moreover, the fool need

controversial and we will be concerned immediately here with its interpretation, it will bear quoting at length:

> The fool hath said in his heart; "there is no such thing as justice"; and sometimes also with his tongue, seriously alleging that: "every man's conservation and contentment being committed to his own care, there could be no reason why every man might not do what he thought conduced thereunto, and therefore also to make or not make, keep or not keep, covenants was not against reason, when it conduced to one's benefit." He does not therein deny that there be covenants and that they are sometimes broken, sometimes kept, and that such breach of them may be called injustice, and the observance of them justice; but he questioneth whether injustice, taking away the fear of God (for the same fool hath said in his heart there is no God), may not sometimes stand with that reason which dictateth to every man his own good; and particularly then, when it conduceth to such a benefit as shall put a man in a condition to neglect, not only the dispraise and revilings, but also the power of other men…[When overtaking a kingdom by force] you may call it injustice, or by what other name you will, yet it can never be against reason, seeing all the voluntary actions of men tend to the benefit of themselves, and those actions are most reasonable that conduce most to their ends. (L xv.4)

The objection here seems to turn on the idea, discussed above, that the same sorts of reasons that lead to conflict and war in the natural condition can be used to show that agents ought rationally to seek peace. Given that a commonwealth is established, and so there is a common power capable of enforcing covenants, the fool notices that there is nothing incoherent in supposing that one break covenant when one can do so with impunity. Not only is such a state of affairs not incoherent, it also seems to be rationally recommended. The fool is supposing that the move out of the natural condition, made rational by reasons of self-interest, has not affected the rationality of pursuing what one takes to be in one's self-interest within the commonwealth. The fool is asking, "what reason do I have for keeping my covenants when not doing so stands to benefit me?" That is, he is asking "why be moral," where being moral involves the making and keeping of covenants in a commonwealth.[36]

---

not deny that the existence of the commonwealth is rational, insofar as it's useful for him to achieve his self-interest.

[36] Both Kavka and Gauthier maintain that the fool's objection applies to the violation of any law of nature, and not merely to the third law. That is, the fool is not merely claiming that the dictates of self-interest outweigh the demands of justice, but that they outweigh any claims made on individuals by any of the moral laws.

Briefly, Hobbes's response to the fool consists of three main claims. First, Hobbes points out that an action is not "reasonably or wisely done" if it is a kind of action that tends to yield unfavorable outcomes. This is so even if "some accident which [the agent] could not expect, arriving, may turn it to his benefit" (L xv.5). Performing an action that typically produces bad results does not become rational merely because, through sheer luck, one happens to avoid these bad consequences on a particular occasion. Second, Hobbes reminds the fool that the whole point of contracting with others is for the sake of self-preservation, and so there can exist no society that willingly allows a known covenant violator into their midst: "He, therefore, that breaketh his covenant, and consequently declareth that he thinks he may with reason do so, cannot be received into any society that unite themselves for peace and defence, but by the error of them that receive him" (ibid.). These sorts of errors are ones "which [the fool] could not foresee nor reckon upon" (ibid.). Finally, because the cost to the agent of being discovered as a violator, and hence being cast out of society, is to be again in the natural condition, and since in such a condition he will "perisheth," Hobbes concludes that, in violating a covenant, "[he has acted] against the reason of his preservation" (ibid.).

The fool's objection and Hobbes's response have received much attention in the secondary literature. In what follows, I will present a few of the interpretive options that have emerged within one strand of Hobbes scholarship. [37] My aim in this section is not to argue in favor of any of these interpretations, and I will present some reasons for thinking that none of

---

[37] I do not intend to give a sampling of all, or even many, of the main *types* of interpretations that have been offered. One particularly prominent strand of interpretation in the early 20th century emerged from the works of A.E. Taylor and Howard Warrender. The *Taylor-Warrender Interpretation* holds that Hobbes's account of moral obligation is independent from his egoistic psychology and that moral obligation depends on the commands of God. See Taylor 1936, 418-20; and Warrender 1957, 98-9, 213, 248. A.P. Martinich is a recent defender of this view (1992, 80; 87-99). Others in this tradition, like Bernard Gert and Rosamond Rhodes, do not accept the theological grounding of the view, but nevertheless maintain that moral obligations arise from categorical imperatives. See Gert 1991, 14-5, 29; Rhodes 1992, 100, 102 n.23; and 2002, 49-60. I have not included a discussion of any of these figures in what follows. This is not because I do not find these views important. I have not included them because I am focused on locating an internalist thread in Hobbes's account, and I think this can be done by paying attention only to those interpretations variously referred to as "traditional," "egoistic," and "prudentialist."

them are entirely adequate. I think each of the interpretations brings out important elements of Hobbes's account of reason and obligation that will be helping in considering what, if any, sort of internalist thesis Hobbes would have accepted. These elements come out when we consider where Hobbes and the fool disagree, what Hobbes is attempting to persuade the fool to believe, and how Hobbes and the fool understand the relationship between the laws of nature and self-interest.

Hampton's Interpretation

According to Jean Hampton, "Hobbes's answer to the fool is remarkable, because it directly contradicts the position taken in the chapters…in which Hobbes appears to adopt the fool's position to explain the failure of contracts in the state of nature."[38] A reasonable suspicion that the other party will not keep his end of the bargain is sufficient to void a covenant of mutual trust in the natural condition because to perform in the presence of such uncertainty would be "to expose [oneself] to prey (which no man is bound to)" (L xiv.5, R 58).[39] According to Hampton, Hobbes claims that one should not perform one's contractual obligations if doing so "means not doing what conduces to one's benefit."[40] This is how Hobbes explains the general failure of covenants in the natural condition. However, the reasoning here does not seem to immediately apply to the fool's objection. While this explains why keeping covenants of mutual trust in the natural condition is not rational, it is not immediately obvious how such reasoning applies to the

---

[38] Hampton 1986, 65; 78-9. For ease of exposition, I here follow Hampton in using the terms 'contract' and 'covenant' interchangeably.

[39] Elsewhere, Hampton notes that one of the fundamental lessons to take from Hobbes's moral philosophy is that morality does not require that individuals "make ourselves the prey of others" (Hampton 1993, 245).

[40] Hampton 1986, 65. This is certainly so if we understand "not doing what conduces to one's benefit" as "doing what conduces to self-harm". However, there is some question whether Hobbes would accept this when "not doing what conduces to one's benefit" is understood as "not doing what would bring about *some* benefit to oneself". There is even a question of whether Hobbes would accept this if "not doing what conduces to one's benefit" is understood as "not doing what would bring about what is *most* beneficial to oneself". For a discussion on these points, see Kavka 1986, 35-64.

fool's objection. This is because, as Hobbes notes, "the question is not of promises mutual where there is no security of performance on either side…but where one party has performed already, or where there is a power to make him perform" (L xv.5). However, Hampton thinks that the fool would have "no problem" with finding covenant keeping in the commonwealth rational since "the sovereign's enforceable laws make it in one's interest to do so."[41] What is primarily of interest for Hampton in the fool's objection is what it tells us about the rationality of keeping covenants in the natural condition.

Hobbes does claim that it is "not against reason" for one to keep covenant in *either* of the two situations to which he takes the fool's objection to apply, namely in a commonwealth or where one party has already performed. So, Hampton suggests, "Hobbes says that it is rational to keep covenants *in the state of nature* if one party has already performed."[42] In addition, Hampton argues that, because it is rational for the second party to perform given first party performance, it is also rational for the first party to perform (given, of course, that the first party is rational and has good reason to think the second party is rational as well). When the first party has good reason to think that her contractual partner is rational, because second party performance is rational given first party performance, Hampton's claim is that first party performance is also rational. She thus finds Hobbes's response to the fool to be puzzling because she sees him as claiming that keeping covenants in the natural condition is rational while recognizing that contracts made in such a condition are void. She is not here committing herself to the view that Hobbes holds that contract making and keeping are therefore widespread in the state of nature. Rather, she is developing an account of the *failure* of contract making and keeping that is not a failure of rationality. It is rational for agents in the state of nature to make

---

[41] Hampton 1986, 64-5.

[42] Ibid., 64.

and keep contracts, but they do not do so because irrational passions, like fear or a desire for glory, get in the way.[43]

We can further see why Hampton thinks that Hobbes's response to the fool directly contradicts his account of the invalidity of covenants of mutual trust in the natural condition by considering the view of rationality that she attributes to Hobbes.[44] First, Hampton takes Hobbes to be an egoist insofar as he holds that only desires can move us to action, and all desires are formed by biological/physiological mechanisms that direct individuals to seek their own pleasure (or avoid pain to themselves).[45] Hampton claims that this is the sense in which Hobbes is a psychological egoist. Hampton, however, denies that she interprets Hobbes as a psychological egoist in the sense that I've been using it here. She maintains that Hobbes distinguished between the objects of desires, and the biological/physiological mechanisms responsible for creating desires. The former are "largely (but not exclusively) self-regarding," while the latter are "exclusively self-interested."[46] Since human beings can have other-regarding desires, Hampton denies that Hobbes held the view that humans can only have self-regarding desires, and so a version of psychological egoism defined as the thesis that humans can only have self-regarding desires cannot be applied to Hobbes. She largely follows Bernard Gert in her argument for this claim.[47] However, she also claims that Hobbes held that all of our desires are generated from "our inherent biological pursuit of pleasure."[48] Hampton wants this view to be distinct from the claim that all voluntary actions are caused by desires that are the agent's own. She often speaks as if the reason for an agent's performing an action is for the sake of his or her own pleasure,

---

[43] Ibid., 63-8.

[44] For this presentation of Hampton's view, I am indebted to Kavka's discussion in 1995, 9-14.

[45] Hampton 1986, 19-24.

[46] Ibid., 24.

[47] See Gert 1967, 508-12.

[48] Hampton 1986, 22.

even if the agent has no conscious or unconscious desire for his or her own pleasure. This seems to rule out the possibility that an individual could desire, for its own sake, some good for another person where this involves a sacrifice of pleasure on the part of the agent. So, there is some basis to say that Hampton attributes to Hobbes a psychology that is egoistic in this way.[49]

There are two other important elements of Hampton's understanding Hobbes's views on rationality. She thinks Hobbes understands decisions under conditions of risk to be rational insofar as such decisions are guided by expected utility calculations. Appealing to the expected utility of various alternatives allows Hampton a way to reconcile human being's fundamental desire for self-preservation with Hobbes's insistence that we do sometimes need to risk our lives. Finally, she holds that Hobbesian rationality consists in case-by-case expected utility calculations, and not in rational rule-following.[50] Rules are useful tools that can tells us what will generally lead to us to our ends, but they have no authority outside of these ends.

It is clear that, for Hampton, if the fool was absolutely certain that (i) he would in fact benefit from the violation of a covenant in the commonwealth and (ii) his violation would not be discovered (and so he wouldn't be harmed either by the sovereign's sanctions or by gaining a reputation as a violator), then Hobbes would say that it is rational for the fool to violate his contract. But, as Kavka notes, the dispute with the fool does not concern cases where the outcomes are known with certainty in advance.[51] Even so, Hampton maintains that the case is not all that different when the outcomes of violation are not known with certainty, but are nevertheless likely to benefit the fool and unlikely to harm him. Given that individuals act only for their own self-interest, and in each case decide which actions are likely to yield the highest expected utility compared to the available alternatives, then if the situation the fool is imagining

---

[49] Kavka 1995, 10.

[50] Hampton 1986, 16-17; 93.

[51] If we interpret the fool's objection in this way, then "there is no substantive point at issue between him and Hobbes" See Kavka 1995, 7.

arises (where the expected utility of violating a covenant is greater than the expected utility of not violating it) then it seems that violation is rational. This is exactly the claim that the fool makes, and, if Hampton is right about Hobbes's view of rational choice under conditions of risk, it should be the claim that Hobbes makes.

It is not, however, the claim that Hobbes makes in his response to the fool. There, he seems to endorse the view that violation is not rational. More surprisingly, he seems to endorse the view that violation is *never* rational. One way of making sense of this is to stress the severity of the harms that would befall the fool if it were discovered that he broke covenant. The idea here is that there is simply too much at stake in violating covenants, and the bad that would befall a violator upon being found out would be so great that, no matter how low the probability of being discovered and no matter how great the likely benefits, the expected utility calculation will always turn out to favor keeping covenants. So, Hobbes's response to the fool is essentially to charge the fool with miscalculating the expected utility of breaking his covenant by either overestimating his ability to go undetected, or by underestimating the severity of the consequences of being caught.

There are quite a few problems with this interpretation, not the least of which is that the view which it claims Hobbes would be defending is, as Kavka puts it, "implausible on its face."[52] For one, there doesn't seem to be any reason to suppose that, when the probability of benefits is high and the probability of being found out is quite low, that the badness of being found out will always weigh expected utility towards non-violation. Such a result could be achieved if, for example, we supposed that the amount of badness upon violation were infinite, and that the goodness of the expected benefits were always some finite quantity. Certainly, the latter claim is available to Hobbes: "*Continual success* in obtaining those things which a man from time to time desireth…is that men call Felicity; I mean *the felicity of this life*" (L vi.58, R 37; emphasis added). But the former claim, that violation of covenants could lead to an infinite

---

[52] Ibid., 12.

amount of badness, does not seem available to him. Hobbes thinks that the principal evil that agents seek to avoid in contracting out of the natural condition is their untimely death. Since felicity consists in the continual satisfaction of an agent's desires, and unhappiness in the frustration of these desires, death doesn't mark an infinite amount of badness. Rather, death removes the possibility of anything being either good or bad. So, it seems that, on the assumption that one will not live an infinite life, death could only ever count as being as bad as the lost quantity of goods one could have achieved if one's life had continued.[53]

The main problem is that fool seems to be presenting us with a question about what it would be rational to do were it to be the case that the probability of benefit was great and the risk of sanction was very minimal. Hobbes cannot say that this is an in principle impossibility. Hampton thinks that the fool would readily accept that it is rational to keep covenant in the commonwealth given the sanctions imposed for violating. But this seems to miss one of the central points of the fool's objection. The only response that seems available to Hobbes is to say that no matter what, if the fool's expected utility calculations make it such that violation of a covenant in the commonwealth is the best option, then the fool has to have made some error in his calculation. In addition to being implausible, it seems this response doesn't fully address the question under consideration.[54]

The position that Hampton develops maintains that Hobbes does not disagree with the fool about violations under conditions of certainty. Hobbes merely says that no violation of covenant occurs under such conditions. On her view, both Hobbes and the fool share a conception of the nature of rational action. Both think that rational action is egoistic, seeks to maximize expected utility, and depends on a case-by-case basis evaluation of actions rather than

---

[53] This may go some way towards explaining those passages where Hobbes seem to recognize that people suffering from great amounts of pain, and who are unable to satisfy any of their desires, may rationally choose death over continued life. Death is only of negative value relative to the goods a continued life could have secured.

[54] Gauthier similarly reads Hobbes's attempt to respond to the fool on a case-by-case, egoist account to be unimpressive and to miss the point. See Gauthier 1989, 548; and 1982, 17.

on a reliance on general rules. This means that Hobbes responds to the fool by correcting the fool's expected utility calculation. The disagreement between Hobbes and the fool on this sort of an account is rather minimal.

Kavka's Interpretation

As Kavka understands it, there is a more robust disagreement between Hobbes and the fool. The main point of their disagreement lies in their different views on the nature of rationality. The fool accepts the view that Hampton ascribed both to him and to Hobbes. Kavka maintains that Hobbes, in fact, held a quite distinct view about rationality. In particular, Hobbes rejects that rationality consists in case-by-case expected utility maximization. Instead, his view of rationality can be cashed out in the following four claims: (i) practical reason is always forward looking from the time of action; (ii) real choices about whether to keep an agreement are made under conditions of uncertainty, "where the probabilities of the various possible outcomes emerging from the available choices are not available";[55] (iii) it is rational to "play it safe" in conditions of uncertainty; (iv) precommitment is rational if "one is likely to do better overall by rigidly following core moral rules than by calculating acceptable risks on particular occasions, because errors and biases in such calculations will tend toward leading you to take excessive risks in particular cases."[56]

The key to Kavka's interpretation is its simultaneous rejection of utility maximization, as seen in claim (iii), and its rejection of a case-by-case rational decision model, as seen in claim (iv). The fool maintains that violation may be rational if particular cases arise where the expected

---

[55] Kavka 1995, 21. It is not clear what Kavka means here when he says that the probabilities of outcomes are "not available." Later on in the same essay, Kavka uses this notion to distinguish "conditions of uncertainty" from "conditions of risk," the latter being cases where the relevant probabilities are known. So, perhaps "unavailable" simply means "unknown." Then, choices about keeping an agreement are made under conditions where we either do not or cannot know the probabilities of the outcomes. But might not we be highly justified in believing what the probabilities are? If so, it is highly misleading to characterize such a condition as one where the probabilities are "not available."

[56] Ibid., 21-2.

utility of such violation is higher than that of performing. Hobbes thinks the fool is mistaken in trusting that the rationality of the violation can be thought of in these terms. Hobbes stresses that the payoff is unlikely, and that the choice is one made under conditions of uncertainty. It is less rational, it is less safe, to follow case-by-case expected utility than to commit to following a set of rules that will tend to procure one's self-interest. In conditions of uncertainty, it is rational to play it safe, and the safer bet here is following the laws of nature.

Kavka thinks that Hobbes's response to the fool not only appeals to a distinct view of rationality from the one the fool presupposes; but he also maintains that Hobbes's response to the fool is, to a certain extent, successful. Kavka notes that there are certain shortcomings of the view of rationality presupposed by the fool. First, he notes that individuals are prone to self-deception regarding their capacities to get away with violations, especially when they stand to obtain something that they desire.[57] Given that mistakes in calculation are likely, it seems safer not to have to rely on case-by-case calculation in order to determine the rationality of an action. Moreover, if an agent happens to do the expected utility calculation correctly on one occasion, and ends up receiving the benefit he thought he would, this will simply reinforce the confidence he has in his abilities to violate without detection. So, says Kavka, "possessing these tendencies, people are likely to err much more on the side of succumbing to temptation and accepting the quick rewards of violation, than on the side of passing up opportunities for gain out of fear of later detection and punishment."[58] The play-it-safe strategy seems to suggest rule-following over case-by-case evaluation. This will not recommend following the rules in every case. Kavka does not think that the play-it-safe strategy amounts to the view that it is always rational to follow the rules encoded in the laws of nature.[59]

---

[57] See also L vi.39, R 31, and L xxvii.16.

[58] Kavka 1995, 22.

[59] Kavka attempts to develop an account that avoids the problem with Gauthier's constrained maximization account, namely that is not "overly rigid" (ibid., 21).

If Kavka's interpretation of Hobbes's response to the fool is successful, it is successful only within a particular range of cases. That is, it only applies to cases where there is in fact widespread uncertainty and where the likelihood of error in expected utility calculations in each case is rather high.[60] Oddly enough, Kavka recognizes this limitation:

> The Hobbesian individualist seeking to defend the rationality of rule-following faces mainly an empirical challenge echoing the remarks of the Foole: is uncertainty really so widespread, and error so likely, that we cannot reliably determine potential violations ahead of time "when… [they] conduced to ones benefit?" (Kavka, 1995, 34).

Yet these are the sorts of empirical conditions to which the fool appeals in his objection to Hobbes, namely, conditions where the expected utility calculation accurately favors violation in a particular case. Given that the play-it-safe strategy is silent with respect to these conditions, it is hard to understand how exactly Kavka thinks he has presented a compelling and successful response to the fool. It seems, instead, that he has simply ignored the fool's objection altogether. So, says Hampton, Kavka's view would look like "in explicable rule-worship" to Hobbes.[61] Both Hampton and Kavka argue that Hobbes does not disagree with the fool about violations under conditions of certainty. They do not agree that Hobbes and the fool share a common conception of rationality. Hobbes and the fool, while agreeing that rationality is in some sense egoistic, disagree insofar as Hobbes thinks that rationality is rule-governed and prioritizes avoiding risks above maximizing expected utility. So, when Hobbes responds to the fool, his response is part of an argument for a rule-governed, play it safe strategy for decisions under conditions of uncertainty. What this amounts to for Hobbes is the view that the laws of nature *can* command an agent not to perform that action with the highest expected utility because following these laws is usually a safer strategy than deciding whether or not to follow them on particular occasions.

---

[60] Hoekstra also raises this objection in Hoekstra 1997, 622.

[61] Hampton 1986, 93.

Hoekstra's Interpretation

Kinch Hoekstra has offered a novel interpretation that bears consideration. [62] First, Hoekstra notes that there are places where there is only limited disagreement between Hobbes and the fool. During the course of his objection, the fool maintains that "all voluntary actions of men tend to the benefit of themselves" (L xv.4). Hoekstra notes that, while this sounds similar to Hobbes's account of the will, Hobbes qualifies the claim in important ways. First, Hobbes recognizes that individuals may be wrong about what they take to be in their own interest, and so he only holds that individuals pursue their *apparent* good. Secondly, Hobbes also recognizes that individuals can fail in their actions to bring about any benefit to themselves. My will may always aim at what is good for myself, but if I am practically inept, then it's not likely my actions will tend to bring about these ends after all.

In addition to this somewhat qualified disagreement, the fool also makes claims about rationality that Hobbes wholeheartedly rejects. The fool claims that, "those actions are most reasonable, that conduce most to their ends" (L xv.4). Hobbes does not accept this claim. He holds that the reasonableness of one's actions depends on the reasonably expected outcomes and not on the actual outcomes. Moreover, the fool seems to presuppose that rationality is understood on a utility maximizing model. But it's not clear that this is the view that Hobbes accepts, as we saw above in Kavka's discussion.[63]

Hoekstra then goes on to note that there are quite a few reasons for taking the fool's objection to hinge on whether or not the fool has publicly declared his view that it may be within reason to violate covenants. Such a fool would, for Hoekstra, be an "Explicit Foole," and his actions would be self-defeating. By declaring that it is in his interest to violate covenants, he thereby makes it less likely that he will be able to enter into those arrangements that could provide him with benefits in the first place. This reading is supported by the many places where

---

[62] Hoekstra 1997, 620-54; 1999, 230-35.

[63] Hoekstra 1997, 622.

the act of public declaration seems vital in the fool passage. For instance, note that, while the fool has said in his heart that there is no such thing as justice, he also says the same thing, only occasionally, with his tongue. So, the fool is someone who, rather stupidly, goes around, *seriously alleging* that it's rational to violate covenants when doing so brings about some advantage to oneself. Also, when Hobbes responds to the fool, he says, "he which *declares* he thinks it reason to deceive those that help him, can in reason expect no other means of safety, than what can be had from his own single power" (L xv.5; emphasis added).

Hoekstra notes that there are a couple of different ways in which the public declaration that violation is reasonable when one stands to gain could be manifested. First, one could be what he calls a "Loud Foole," who states openly and publicly that injustice is reasonable. Alternately, one could be a "Flagrant Foole," who violates covenants in such a way that is so blatant and obvious that his actions are like an open and public declaration that he thinks it is within reason to violate.[64] Both Loud and Flagrant Fooles are distinguished from Silent Fooles, who neither publicly declare they think it within reason to violate, nor who violate in such a way that their actions could be considered as a public declaration. The Silent Foole is the Foole who violates without being discovered, or is only discovered when it is too late for anyone else to do anything about it.

Hoekstra argues that, if we are to take Hobbes's response to the fool seriously, then we must understand him to be addressing only the Explicit Foole, of either kind, and not addressing the Silent Foole at all. If Hobbes were to respond to the Silent Foole, then his response suffers from the sort of weakness we saw in response to Hampton's interpretation. It's just not plausible to suppose that there will never be an instance where an individual cannot accurately judge that violation of a covenant would be more in an agent's self-interest than non-violation. If we

---

[64] The Flagrant Foole is someone who acts in such a way that their behavior can reasonably be interpreted as the behavior of someone who thinks it is sometimes reasonable to violate. So, if Hobbes does think there is such a thing as a Flagrant Foole, we do not need to interpret him as holding the view, later adopted by William Wollaston, that the act is literally a form of declaration.

understand the fool to be a Silent Foole, then Hobbes's response must show that it is in principle never in the agent's interest to violate a covenant, and this, it seems, is implausible. While Hobbes does seem to indicate that such violations are unlikely to be in an agent's interest, given the likelihood of mistakes in the calculations, this is not sufficient to show that such violations will always be irrational.

However, if we interpret the fool as an Explicit Foole, then not only do we get a genuine disagreement between the fool and Hobbes, but Hobbes's response to the fool becomes more plausible. Explicit Fooles are found out, and so we know that future cooperation with them is to be avoided. These are not the sorts of contracting partners that rational and self-interested individuals will allow to be part of their society. It is genuinely a mistake to allow Explicit Fooles into a society that is uniting itself for mutual protection, and these Fooles openly advocate for the performance of those actions that undermine the peace necessary for that protection. Hobbes's reply, if we understand it to target the Explicit Foole only, is then the in principle claim that being an Explicit Foole is never rational, and that this is because Explicit Foolishness can never be tolerated in society.

Silent Foolishness, however, is of less concern. Where an individual can get away with violations, and such violations can be reasonably thought to be in an individual's interest, then there is no in principle reason why such violations are not reasonable. So, according to Hoekstra's interpretation, Hobbes is silent about the Silent Foole, but, since they share overlapping views of the nature of rationality, it seems Hobbes would have nothing *to* say to the Silent Foole (other than to recommend he double check his calculations before proceeding). That Hobbes gives this sort of tacit approval to the Silent Foole would be, I think, a problematic consequence of his discussion of the fool. I take it that Hobbes is there trying to offer an account of how rational self-interest can be squared with the dictates of morality. If we understand the fool's passage as Hoekstra recommends, then we will understand Hobbes as holding the view that the dictates of self-interest, whenever they come into conflict with the dictates of morality, are always to be given priority. This means that individuals should not choose to act for the sake

of, say, keeping a promise, if by not keeping it, they are able to better secure their own self-interest. I will attempt to show below why I think this is a mistaken interpretation of Hobbes.

Another problem facing Hoekstra's interpretation is that it seems to straightforwardly identify the fool with the social agitator.[65] While it's true that Hobbes was particularly concerned to suppress social upheaval, and this suppression often took the form of discouraging any opposition to sovereign power, it's important to keep in mind that the consequences of allowing such dissent, Hobbes thought, was civil war. Dissent and opposition to civil power encouraged the breakdown of that power, and the possible overthrowing of the sovereign, and this, in turn, leads to civil war. Nowhere in Hobbes's response to the fool is he concerned with the result of civil war. Perhaps he is indirectly concerned with it insofar as he warns others not to allow a known violator into their midst, and so he might be thinking that this is what is required in order to avoid civil war. But there arises here the further question about why the fool, who's interested in *circumstantially* advantageous violations, would be a threat for encouraging civil dissolution. The fool is taking advantage of the fact that others are cooperating. He is not inciting aggression towards the state, he is, rather, noting that he can gain something that is in his interest by, occasionally, violating covenants within the cooperative system. The fool is the last person to incite social upheaval given that his self-interested gains depend on the cooperation of others. There is, then, no threat of civil war, and so no reason to identify the fool with the civil agitator.

Finally, as I noted at the beginning of this discussion of Hoekstra, it seems as though the fool and Hobbes are relying on distinct notions of rationality. Since Hoekstra wants there to be a genuine and interesting disagreement between the fool and Hobbes, and we cannot get this on interpretations like Hampton's, he thinks we need to move to the Silent/Explicit model. However, we do get an interesting and genuine disagreement between Hobbes and the fool if we note that they are appealing to distinct notions of practical rationality. Hobbes then would merely be arguing that the fool is appealing to the wrong notion of practical reason, and so is generating

---

[65] This objection comes from Hayes 1990.

the wrong results. This suggestion is in line with Kavka, and, as noted above, Hoekstra doesn't think this model is successful. This is because he, rightly, argues that Kavka's own presentation of the view is unclear at best, and self-defeating at worst. However, simply because Kavka's own attempt to ground Hobbes's response in a rule-following, non-maximizing practical rationality, fails doesn't mean that all such attempts will fail.

Hoekstra, like Hampton and Kavka, thinks there is a sense in which the fool and Hobbes are in complete agreement. Hobbes does not disagree with the Silent Foole fool about either *ex ante* or *ex post* violations. Hoekstra favors reading rationality as Hampton suggests, and so holds that rationality is egoistic, seeks to maximize utility, and governs actions on a case-by-case basis. What is important about Hobbes's response to the fool, however, is that it is only intended to address the Explicit Foole, i.e. the fool who in some sense declares he thinks it's reasonable to violate. Since Hobbes has nothing to say to the Silent Foole, he thinks that it is reasonable to violate the laws of nature in certain circumstances. What this means is that the laws of nature can prescribe actions that an agent's self-interest does not prescribe, and when such divergence of prescriptions occur, self-interest overrides the laws.

Again, my primary intention in these previous sections has been to recount a few of the main interpretations of Hobbes's exchange with the fool. The one claim that is consistent across all three accounts provided here is that there are certain conditions that, in principle, would allow Hobbes to say that violating a covenant, thereby contravening one's moral duty, is rational: these are conditions where the potential violator knows with certainty that she stands to gain and that she will not suffer sanctions. However, it is important for Hobbes's theory that we recognize these conditions never obtain. It is in light of the uncertainty and the great risk that Hobbes recommends keeping our agreements. This does not seem like a promising start to an internalist moral theory that recognizes a necessary connection between moral obligations (defined by Hobbes in terms of keeping valid covenants) and motives or reasons for acting (defined by Hobbes in terms of what is good for an agent. However, in the next chapter I will argue that Hobbes does accept some version of internalism. Before moving on to the arguments for this

claim, I end this chapter with a discussion of the nature of moral philosophy (broadly construed) according to Hobbes.

Hobbes's Moral Terms

Before we can begin to see the sense in which Hobbes might have accepted a version of internalism, we must first identify those terms that Hobbes took to be involved in the science of moral philosophy. I will argue that moral terms for Hobbes essentially involve the passions of human beings *qua* socially related beings. I argue for this point by considering how moral philosophy, ethics, and politics are related to each other and to other sciences.[66] There are a few passages that are important in this regard. The first comes from the *De Corpore* division and classification of the sciences:

> After physics we must come to moral philosophy; in which we are to consider the motions of the mind, namely, appetite, aversion, love, benevolence, hope, fear, anger, emulation, envy, &c.; what causes they have, and of what they be causes. And the reason why these are to be considered after physics is, that they have their causes in sense and imagination, which are the subject of *physical* contemplation…And because all appearances of things to sense is determined, and made to be of such and such quality and quantity by compounded motions, every one of which has a certain degree of velocity…therefore, in the first place, we are to search out the ways of motion simply (in which geometry consists); next the ways of such generated motions as are manifest; and, lastly, the ways of internal and invisible motions (which is the enquiry of natural philosophers). (DC 6.6, 72-3)

Hobbes here takes moral philosophy to be a branch of physics. At a minimum, he seems to hold that one's knowledge of moral philosophy depends upon one's prior knowledge of geometry and natural philosophy, including physics. His explanation for this is that sense and imagination are natural phenomena. In particular, he thinks both sense and imagination (which is a fairly broad term for Hobbes) are kinds of motions. The various passions to which people are susceptible, like appetites, hopes, fears, etc., are also particular kinds of motions. Moral philosophy is, then,

---

[66] My thinking on these issues has been greatly influenced by Tom Sorell. See Sorell 1996, 45-61.

the science that deals with particular kinds of motions, their causes and their effects, and these motions are the various psychological states that Hobbes calls passions.

There does not appear to be a difference between the *De Corpore* characterization of moral philosophy, and the *Leviathan* characterization of *ethics*. In *Leviathan* Ch. ix, Hobbes provides a table of the division of the sciences, and on this table he characterizes ethics as the study of the "consequences from the *passions* of men" (L. ix). Yet, when we turn to *De Homine*, ethics seems to be treated differently than it was in *Leviathan*. In *De Homine*, Hobbes characterizes physics as an a posteriori science that depends on some a priori basis.

> Since one cannot proceed in reasoning about natural things that are brought about by motion from the effects to the causes without a knowledge of those things that follow from that kind of motion; and since one cannot proceed to the consequences of motions without a knowledge of quantity, which is geometry; nothing can be demonstrated by physics without something also being demonstrated a priori (DH x.5).

If ethics were a branch of physics, then we would expect ethics to be an a posteriori science that depends on some a priori demonstrations as well. This would resonate with the *De Corpore* passage cited above. However, this is not how Hobbes characterizes ethics in *De Homine*:

> Finally, politics and ethics (that is, the sciences of *just* and *unjust*, of *equity* and *inequity*) can be demonstrated *a priori*; because we ourselves make the principles – that is, the causes of justice (namely laws and covenants) – whereby it is known what *justice* and *equity*...are. For before covenants and laws were drawn up, neither justice nor injustice, neither public good nor public evil, was natural among men any more than among beasts. (ibid.)[67]

This passage suggests that politics and ethics are distinct from physics. This is partly in keeping with the division of the sciences in *Leviathan.* In the Ch. ix table, politics, or what Hobbes there calls *civil philosophy*, is a distinct field from physics insofar as the former deals with the properties of artificial, as opposed to natural, bodies. However, *Leviathan* Ch. ix does list ethics as a branch of physics, as we've seen. Hobbes also lists "the science of just and unjust," as

---

[67] For a helpful discussion of Hobbes's account of a priori demonstration see Gauthier, 1997.

another branch of physics, and he describes this as the science of the "consequences from speech…in contracting" (L ix). So, the *Leviathan* table separates *ethics* from *the science of just and unjust*, and treats both of these as branches of physics. Both claims are in direct contradiction to the *De Homine* passage above.

It isn't until the end of Ch. xv and the discussion of the laws of nature in *Leviathan* that Hobbes presents a rather different characterization of moral philosophy than what we find in *De Corpore*:

> For moral philosophy is nothing else but the science of what is good and evil in the conversation and society of mankind. Good and evil are names that signify our appetites and aversions, which in different tempers, customs, and doctrines of men are different…And therefore so long a man is in the condition of mere nature…as private appetite is the measure of good and evil; and, consequently, all men agree on this, that peace is good; and therefore also the way or means of peace (which, as I have shewed before, are *justice, gratitude, modesty, equity, mercy*, and the rest of the laws of nature) are good (that is to say, *moral virtues*), and their contrary vices, evil…Now the science of virtue and vice is moral philosophy; and therefore the true doctrine of the laws of nature is the true moral philosophy. (L xv.40, R 77)

Unlike what we see in *De Corpore*, Hobbes in *Leviathan* does not understand moral philosophy to be primarily concerned with the passions, and so it makes sense to think that, in *Leviathan* anyway, Hobbes maintains that ethics and moral philosophy are distinct sciences.[68] This claim might be concealed by the fact that the *Leviathan* characterization does make some reference to the passions; namely, the passage says that moral philosophy deals with good and evil and these, in turn, are understood in terms of appetites and aversions, that is, in terms of passions. However, Hobbes does not say that moral philosophy is concerned with appetites and aversions as such, but that it is concerned with appetites and aversions "in the conversation and society of mankind." Hobbes seems to be saying that moral philosophy, unlike ethics, is not restricted to

---

[68] Peacock say, "obligation (a consequence of speech in contracting) belongs to the '*Science* of Just and Unjust' which Hobbes separates from ethics. It is therefore of *justice*, not morality, that we should write when considering obligation." (2010, 445). He conflates ethics and moral philosophy in a way that overlooks Hobbes's discussion at L xv.40. Hobbes there says morality is the science of the laws of nature, including the law defining justice, which suggests that to write of justice is to write of morality.

understanding the causes and effects of the passions of individuals *qua* individuals. Rather, moral philosophy is the science of those passions of individuals insofar as they engage in social interactions.[69] Moral philosophy is "other regarding" in the very minimal sense that genuine moral terms only apply to the passions involved in the interactions between individuals.

It's not entirely clear what to make of the differences between the *De Corpore* passages, which seem to represent Hobbes's most developed account of the division and classifications of the sciences, and the passages in *Leviathan*. What is clear, however, is that there are two, seemingly independent, sets of terms that may be considered moral terms by Hobbes's categorization.[70] The first set involves terms that will fall roughly under moral psychology. These we understand insofar as we are doing physics, and they include appetite, aversion, love, hate, hope, fear, contempt, benevolence, and a few others that Hobbes takes to be the *foundational passions*. I say they are foundational because Hobbes says that, even though there are "an almost infinite number of passions" he takes himself to be justified in stopping the discussion where he does "since none their be that are not related to some one of those that we have described" (DH xii.12). So, the passions that Hobbes describes are those in terms of which all others will be understood. The second set of terms includes terms like law of nature, justice, gratitude, virtue, vice, and so on.

This bifurcation between the central terms of Hobbes's moral philosophy (broadly construed) gains further support when we turn to *De Homine*, which was published after both *De Corpore* and *Leviathan*. In *De Homine*, Hobbes says:

---

[69] "Hobbes always using the word 'moral' with regard to those virtues that lead to peace, and neglects to use it with regard to those virtues that lead to individual preservation" (Gert 1967, 511). I endorse the spirit of Gert's claim here, even though there are cases (*DeCorpore*) where this use of 'moral' is unclear.

[70] Some commentators have argued that this dualism of language arises because Hobbes's psychology is not dependent on his mechanistic physics. Gert, for instance, argues that even though Hobbes thought he could develop a psychology from physics, "he made no serious attempt to do so," and that Hobbes recognizes that we can arrive psychological truths through introspection, but we cannot arrive at truths about motion via introspection. See Gert 1967, 503-4.

> Moreover, that moral virtue, that we can truly measure by civil laws, which is different in different states, is justice and equity; that moral virtue is which we measure purely by the natural laws is only charity. Furthermore, all moral virtue is contained in these two. However, the other three virtues (except for justice) that are called cardinal – courage, prudence, and temperance – are not virtues of citizens as citizens, but as men, for these virtues are useful not so much to the state as they are to those individual men who have them. (DHxiii.9)

For present purposes, I want to focus on the distinction that Hobbes draws between "citizens as citizens" and "citizens…as men." This distinction, I suggested, was already present in *Leviathan*, and it could be seen when considering the difference between ethics and moral philosophy. Now, in *De Homine*, Hobbes appears to make the same distinction, only he makes it much more clearly, and uses much more helpful terminology. He here distinguishes between the virtues of individuals and the moral virtues. Hobbes also says:

> Nevertheless, what is to be understood about men insofar as they are men, is not applicable insofar as they are citizens; for those who are outside of a state are not obliged to follow another's opinion, while those in a state are obliged by covenants. Whence it is to be understood that they, who consider men by themselves and as though they existed outside of civil society, can have no moral science because they lack any certain standard against which virtue and vice can be judged and defined. (DH xiii.8)

So, there is reason for thinking that moral philosophy is a matter of understanding the nature of the moral virtues as they arise from the laws of nature. If we are to see whether Hobbes accepts a version of internalism, then we would ask what Hobbes takes the relationship between justice, gratitude, mercy, etc. and the motives and reasons of agents to be. In what follows, I will restrict my attention to the virtue of justice (since it is crucially what is at stake in the fool's objection). However, as we will see, there doesn't seem to be any reason for denying the internalist relationship to hold between the states of the agent and other virtues. That is, we will see that Hobbes accepts a broad version of internalism, ranging over all the moral virtues. In addition, since there is a fundamental relationship between justice on the one hand, and obligation and moral ought claims on the other, the broad version of internalism will apply equally to obligation and moral ought claims as to the other moral virtue terms.

CHAPTER 2

INTERNALISM, OBLIGATION, AND THE FOOL'S PRACTICAL IRRATIONALITY

Introduction

Moral internalism asserts that there is a necessary connection between reasons for action (understood as either explanatory or justifying reasons) and *moral* or *ethical* judgments or properties. One way in which we could approach the issue of understanding the sense in which Hobbes could have accepted a version of internalism is to restrict our attention to his understanding of *moral* or *ethical* terms, concepts or properties.

Darwall, however, has a different approach. In discussing the development of modern morality, as distinct from ancient or medieval morality, Darwall notes that what is distinctive about it is its insistence that morality consists in binding demands on all rational agents, regardless of their own self-interest. It is in this connection that Darwall notes the pressing need to answer the more fundamental questions: "What is it for anything to be binding? What is bindingness, obligation, or, as moral philosophers are inclined to say these days, *normativity* itself? In what does an *ought to do* consist?"[1] Elsewhere, Darwall notes that this "ought to do" will always be of an ethical nature: "Whether there are reasons to act is a normative and thus, in the broad sense, an ethical matter. It concerns what the agent rationally *ought* to do."[2] So, when he describes the sort of internalism he credits Hobbes for initiating, Darwall says it's the view that "[a]n agent ought to do something if, and only if, she would be moved to do it were her empirical beliefs about her practical situation error-free."[3] He does *not* limit the view to considerations about what one *ought morally* to do, and so we

---

[1] Darwall 1995, 2. As we will see, Darwall's question about what obligations *consist in* leads to some difficulty in understanding his interpretation of Hobbes.

[2] Darwall 1992, 156.

[3] Darwall 2000, 317. The sense in which this is a version of internalism will be made clear in what follows.

can conclude from this that the sort of internalism he has in mind is meant to be true of practical ought statements *as such*.

As Darwall understands him, then, Hobbes maintains that "[e]thical thoughts – concerning good and ought – are those an agent has in *deliberation*, as her desires and aversions are expressed *seriatim*."[4] I will specify the relationship between the practicality of ethical thought and the serial expression of desires and aversions in the discussion of the nature of deliberation and the will below. The significant thing to note here is that Darwall understands the terms 'good' and 'ought' to both be equally a part of *ethics*, and so both will give rise to internalist analyses.[5] There is a sense in which this claim is right, namely, that we can find versions of internalism in Hobbes's account of both terms. However, Darwall here runs roughshod over Hobbes's distinction between ethics and moral philosophy.[6] Moreover, as I argue below, the version of internalism that can be seen in Hobbes's account of 'good' is different from the sort of internalism that can be seen in Hobbes's account of 'ought' and the moral terms generally, though the two are related.[7]

Internalism with Respect to 'Good'

There is one rather clear, albeit trivial, sense in which Hobbes would accept a version of motive internalism narrowly interpreted about goodness. Hobbes says,

> But whatsoever is the object of any man's appetite or desire that is it which he for his part calleth *good*; and the object of his hate and aversion *evil*; and of his

---

[4] Ibid., 332. I criticize this characterization insofar as it leaves out the essential element of moral philosophy, namely the desires and aversion *of citizens*.

[5] It seems as though Darwall here is asserting that both 'good' and 'ought' are *normative* terms. While there may be a relationship between them, it seems odd to simply lump them together as being equally normative, when many philosophers assert that there is a difference between the evaluative and normative terms. See Wedgewood 2009. However, even if one does not accept this evaluative/normative distinction, it should be clear from the previous chapter that Hobbes distinguished ethics from moral philosophy.

[6] It's clear that Hobbes held there was distinction between ethics and moral philosophy, though he is not always clear in his statements about what exactly that distinction involves.

[7] In terms of his account, however, Darwall's treating statements of 'good' and 'ought' as both normative does not make much of a difference. I point to this merely to show one respect in which Darwall's interpretation might be improved to be more textually adequate.

> contempt, *vile* and *inconsiderable*. For these words of good, evil, and contemptible are ever used with relation to the person that useth them, there being nothing simply and absolutely so, nor any common rule of good and evil to be taken from the nature of the objects themselves. (L vi.7, R 25)

An agent will call good whatever objects she happens to desire, and as I understand him here, Hobbes is saying that all there is to something's being good for S is for S to desire it. Desire and aversion are species of endeavors, for Hobbes. Endeavors are small internal motions that begin in imagination, which is itself a kind of motion.[8] Hobbes thinks that external objects act upon the sense organs to give rise to various kinds of internal motions, and these motions in their various operations are all that sense, conception, and imagination consist in. Sense, conception, and imagination are all motions that occur within the head, but this motion does not remain within the head alone: "so when the action of the same object is continued from the eyes, ears, and other organs to the heart, the effect there is nothing but motion or endeavour, which consisteth in appetite or aversion, to or from the object moving" (L vi.9, R 26). Here we see that directedness is a property of endeavors, and this property serves to distinguish appetites from aversions. When the motions are directed towards the object that caused them, they are desires or appetites, and the object that caused them is called good. When these motions are directed away from the object that caused them, they are aversions, and the object is called evil.[9]

By distinguishing appetites and aversions in terms of the motion either toward or away from an object, Hobbes is not saying that every appetite and aversion will be manifested in an individual's overt, external behavior. One might attempt to claim that this is his intended meaning by noting that the discussion of appetites and aversions here takes place within a discussion of "*animal motion*, otherwise called *voluntary motion*, as to *go*, to *speak*, to *move* any of our limbs, in such manner as is first fancied in our minds" (L vi.1, R 23). However, Hobbes makes it clear that the internal motions

---

[8] L ii.2 See Gert 1996, 157-8.

[9] It is interesting to note that what Hobbes says here is that 'good' and 'evil' are what individuals *call* the objects of their desires and aversions, respectively. He does not say that good are evil are to be *identified* with the objects of our desires and aversions. This allows some room for denying that Hobbes was a metaethical subjectivist. This view seems to amount to an error-theory. See Darwall 2000.

still have directedness even when they fail to be manifested in external actions. Hobbes maintains that the same object can give rise, in the same person, to a series of alternating desires and aversions:

> When in the mind of man appetites and aversions, hopes and fears, concerning one and the same thing arise alternately, and diverse good and evil consequences of the doing or omitting the thing propounded come successively into our thoughts, so that sometimes we have an appetite to it, sometimes an aversion from it, sometimes hope to be able to do it, sometimes despair or fear to attempt it, the whole sum of desires, aversions, hopes and fears, continued till the thing be either done or thought impossible, is that we call DELIBERATION. (L vi.49, R 33)

In deliberation, desires and aversions present themselves in succession, and so there is also an alternating series of motions towards and away from some object. If we understood this motion as the motion of external action, then for all processes of deliberation we would expect to find continually shifting external behaviors. Hobbes knows that this is a ridiculous suggestion, and so defines the will as the last appetite or aversion in deliberation (L vi.53, R 33). The will is the endeavor that takes, so to speak, and so all other endeavors in deliberation cannot be manifested in overt, external behavior.

Rather, Hobbes maintains that the directedness of desires and aversions is a property of whatever motion connects the motion around the heart with the motions necessary for external action. There is some textual support for this claim in *Human Nature* where Hobbes defines the motive power of the mind (as opposed to the motive power of the body, which is identified as strength) as "that by which the mind giveth animal motion to that body wherein it existeth" (HN vi.9). This motion is such that, if there are no competing desires or aversions, hence no deliberation, external action towards or away from some object will result (given that no physical impediments are present). In other words, we can say that the directedness of desires and aversions, either towards or away from, is partly causally responsible for the direction towards or away from some object in an individual's external actions. In *Human Nature*, Hobbes refers to this feature of endeavors as "a *solicitation or provocation* either to draw near to the thing that pleaseth, or to retire from the thing that displeaseth" (HN vii.2, emphasis added). The directedness of the endeavor is what determines the direction of the external action, other things being equal.

If this is correct, then we can begin to see a sense in which Hobbes could have accepted a certain version of motive internalism. On Hobbes's view, when an agent, S, judges that some object, X, is good, it follows analytically that S desires X. It follows analytically that if S desires X, then there is in S a motion towards X. Now, if S's desire for X is the last, or only, endeavor in deliberation, so that S's desire for X is her will, so long as she is not physically constrained by external impediments, then S will manifest external actions to obtain X. When S desires X, other things being equal, she wills to obtain X. When other things are not equal, that is, when S is in deliberation with respect to X, it is still true that that the directedness of this endeavor determines the direction of her external actions. If we stipulate that motives are the sorts of things that determine the course of external actions in this sort of way, then S will have *some* motive to obtain X. So, Hobbes would accept a weak version of motive internalism: in virtue of her judging that X is good for her, S would have *at least some* motive to obtain X. This, indeed, is a tautology for Hobbes. It is equally a tautology to claim that, in the absence of alternating desires and aversions, and when S is not physically constrained by external impediments, S's desire for X is a *sufficient* motive. This is because under such conditions, her desire for X is just her will, and her act of willing by definition just is her having a sufficient motive to act.

Moreover, there is textual reason for thinking that desires and aversions are motives on Hobbes's view.[10] First, as mentioned above, Hobbes describes the motive power of the mind as that motion which underlies the overt behavior towards or away from the objects of our desires or aversions. Additionally, he says in connection with this that, "the acts hereof [those motions] are our affections and passions" (HN vi.9). Also, in *Leviathan* Ch. xi, Hobbes gives an account of the various actions that tend to follow upon agents having certain kinds of desires. He argues that pusillanimity, or the overvaluing of trifles, tends towards irresolution:

> For after men have been in deliberation till the time of action approach, if it be not then manifest what is best to be done, *it is a sign the difference of motives,*

[10] Some commentators maintain that desires (and likewise, presumably, aversions) are the *only* things that move us on Hobbes's account of voluntary action. See Hampton, 1986 pp.20-22.

> *the one way and the other*, are not great; therefore not to resolve then is to lose the occasion by weighing of trifles; which is pusillanimity. (L xi.14; emphasis added)

We need not accept what Hobbes has to say here about pusillanimity and inaction, though it seems plausible enough. What's important for present purposes is to note that the directionally distinguished elements of deliberation, desires and aversions, are here referred to as motives. Finally, and perhaps most significantly, is what Hobbes says when he identifies the will with the last endeavor in deliberation: the will is "the last appetite or aversion [in deliberation] *immediately adhering to the action, or to the omission thereof*" (L vi.53, R. 33; emphasis added). With this textual support, the supposition that Hobbes held a version of internalism about judgments about goodness is, I take it, fairly well established.

It should be noted that this is not merely a version of appraiser internalism. An appraiser internalism about good would maintain that, necessarily, if S genuinely judges that X is good, then S has some motive towards X. Hobbes does accept that this is the case: S's having a motive to X follows analytically from S's judging that X is good. However, S does not inherit her motivational state from the judgment that X is good. Her motivational state arises from the interaction between an external object, her sense organs, her previous experiences constituting memory, and her occurrent judgments. S would be in the motivational state she was in with respect to her desire for X even if S had never judged that X was good. I am here assuming a distinction between *S desires X* and *S judges that X is good for her*. If the former can be true while the latter is false, it is still true that S has some motive towards X. Now, Hobbes might say that, for human beings, a desire for X is always accompanied by a judgment or opinion that X is good, but in these cases the motivation is not a result of these accompanying judgments and opinions. Also, the motivational state underlying judgments involving 'good' are identical to the motivational states underlying the actions of nonhuman animals that lack the capacity for naming and making judgments. In his discussion of the nature of deliberation, Hobbes says, "[this] alternate succession of appetites, aversions, hopes, and fears is no less in other living creatures than in man; and therefore beasts also deliberate" (L vi. 51, R 33). Hobbes denies that nonhuman animals are capable of judgment, at least where this involves

understanding the logical consequences of "the names of things [in] affirmations, negations, and other forms of speech" (L ii.10).[11] Yet, the necessary connection between desire and motivation does not presuppose that an animal has any understanding or capacity for judgment whatsoever. So, it isn't plausible to suggest that Hobbes's internalism about good is *only* a thesis about an individual's judgments involving 'good'.[12]

If it's possible to separate S's desire for X from her judgment that X is good, then S would have the same motivational states with respect to X were she never to judge that X is good. Since goodness is identical with the objects of an individual's desires, for Hobbes, it follows that, necessarily, if X is good for S, then S has some motive towards X.[13] In other words, Hobbes would also accept a version of existence internalism with respect to goodness.

I think it is unlikely that Hobbes would take this sort of internalism to be of the non-constitutive variety. Non-constitutive versions of internalism hold that there is a necessary connection between an agent having cognitive contact with some normative, or moral, fact and an agent being in some motivational state.[14] The non-constitutive interpretation can be ruled out on the same grounds

---

[11] Hobbes recognizes no difference in kind with respect to the imagination and understanding between human and non-human animals. The difference is one of degree.

[12] I think those who offer a projectivist interpretation of Hobbes's theory of value can accept this account. The projectivist denies that goodness is identified with the objects of an individual's desires for Hobbes. Instead, the projectivist maintains that goodness appears to individuals as if it were an objective normative property of objects (or states of affairs). This appearance is caused by the individual's desires. So, the projectivist maintains that when an individual desires X, her desire for X makes it appear to her as if X had the property of being good. On this account, then, there is no room for separating an individual's desires for X and her awareness that X is good for her. Nevertheless, since the projectivist holds that there is no such property as an objective, normative property of goodness, it seems clear that the motivational states of the agent depend on her desires, and not on her perceiving the objects of her awareness as good. See Darwall 2000.

[13] The projectivist would likely deny that an individual's judgment that X is good for her can be separated out from her desire for X in this way. The account is further complicated since the projectivist is committed to distinguishing judgments that X is good from the appearance of X as good (just as they must distinguish the judgment that X is red from the appearance of X as red). See Darwall, 2000.

[14] Darwall does not address whether the projectivist interpretation commits Hobbes to accepting a version of non-constitutive version of existence internalism. There is a sense in which the projectivist cannot accept this version of internalism: insofar as they deny that there are normative properties for agents to be in cognitive contact with, they also deny non-constitutive internalism so understood. However, it might be worth considering whether a distinct account of non-constitutive internalism might

that allowed us to rule out the possibility that this version of internalism was a species of appraiser internalism: such a view presupposes cognitive capacities that Hobbes would not ascribe to nonhuman animals, and there doesn't seem to be any good reason for holding that the cognitive contact with the normative, or moral, facts is necessary to understand the relationship in the human case either. The sort of internalism about good that Hobbes would accept is constitutive existence motive internalism. Hobbes holds that all there is to something's *being* good is that agents are in certain motivational states with respect to it. Goodness is not a *sui generis* property and has no existence independent from the desires of individuals: "Nor is there any such thing as absolute goodness, considered without relation" (HN vii.3).

What should be clear from this account is that this version of internalism will not by itself be useful in understanding Hobbes's answer to the question "why be moral?" This version of internalism is, as we have seen, tautological. The only way in which it can help us address the question "why be moral?" is if we presuppose that individuals call good the ends of those actions that are considered moral. And while Hobbes does say that, "all men agree on this, that peace is good; and therefore also the way or means of peace…are good," this claim requires additional support (L xv.40, R 77). In the next section, I will discuss this additional support, and will also discuss the sort of internalism that Hobbes would take to apply to 'ought' claims and how this relates to internalism about good.

Empirical Naturalist Internalism

It should be clear from what was said above that Hobbes does not think that certain forms of amoralism arise with respect to goodness and motivation. It cannot be the case that X is good for S, while S has no motive to act for the sake of X. There may be competing, stronger motives, but all of these will be understood as being of the same kind and involving others goods for S. But in virtue of S's desiring X, which necessarily follows from the fact that X is good for S, S will have at least some

---

be constructed on the basis of an individual's perception of goodness *as if it were an objective, normative* property. Such a discussion would be too much of a digression from main issues with which I am here concerned.

motive. It is not clear that Hobbes would deny the possibility of the amoralist with respect to judgments about what one ought to do. For Hobbes, it is analytic that one ought to do whatever it is that one has contracted to do (L xiv.7, R 59).[15] However, as Darwall points out, "Hobbes recognizes a sense in which it is far from analytic that people ought to keep covenant."[16] This leads us to recognize two distinct senses of 'ought' in Hobbes's moral theory. First, there is the sense of 'ought' that follows analytically from the fact that one has made a contract. Second, it is also true that individuals ought to follow the dictates of right reason. I will, following Darwall, call these senses of ought the obligation 'ought' and the law 'ought', respectively, and I will refer to them as ought_O and ought_L. The two senses come apart when we consider that one may have entered into a contract (and so it's true of her that she ought_O to keep it), but she might still wonder if she has reason to keep it (that is, whether or not she ought_L to keep it). Indeed, this seems to be precisely the situation that the fool presents us with.[17]

A few remarks must be made at this point before moving on to consider the relationship between obligations, ought_O and ought_L claims, and their relationship to internalism about goodness. Darwall maintains that Hobbes held a version of empirical naturalist internalism, and part of my aim in this chapter is to develop this account as far as possible. Darwall's discussion is problematic insofar as he never clearly develops how this view falls within the general taxonomy of internalist views that he provides. Moreover, as we'll see, Darwall doesn't think that all empirical naturalists even accepted versions of internalism. According to the empirical naturalist view in general, agents are understood to have no epistemological faculties besides those involved in empirical inquiry (including a capacity for

---

[15] In what follows, I will continue to use 'contract' and 'covenant' interchangeably. There is a distinction between them for Hobbes. However, since Hobbes holds that one ought to do whatever one has laid down his right to do, and both contracts and covenants involve such a laying down of a right, it follows that one ought to do both what one has contracted to do and what one has covenanted to do. So, the technical distinction between them will not make a difference for the following discussion.

[16] Darwall 1995, 56.

[17] The fool may recognize that he has some reason to keep his covenants in virtue of the fact that he made the agreement. He would then be best understood as asking if he has an all things considered reason to keep covenant.

demonstrative, or deductive, reasoning). The view holds that moral truths can be explained in fully naturalistic terms, and so denies that moral philosophy needs to appeal to any non-natural facts.

Now, the way Darwall presents the empirical naturalist internalist view, it seems like he takes the view to be a version of constitutive existence motive internalism. To restrict our attention to obligation, necessarily, if an agent S is obligated to Φ, then, if her beliefs about her practical situation were accurate, S would have some motivation to Φ in virtue of a constitutive relationship between motivation and obligation. The constitutive relationship is one such that being obligated to act is to be understood in terms of motives that are discovered through the use of theoretical reason. Darwall says that central commitment of empirical naturalist internalism is that "*obligation consists in motives raised through the use of theoretical reason*."[18] To be obligated, then, just is, at least in part, to have (under certain conditions) certain kinds of motives and these are the motives that would arise from the correct use of theoretical reason. As Darwall puts it, "An agent ought to do something if, and only if, she would be moved to do it were her empirical beliefs about her practical situation error-free."[19]

There are a couple of issues that need to be clarified regarding this account of internalism. First, it is clear that obligation is not to be understood in terms of motives that an agent necessarily actually presently has, though she may in fact have them. As Darwall seems to understand it, the obligation to Φ is, at least in part, *constituted by* motives that an agent would have were she to reason correctly about her practical situation.[20] When she does not reason correctly about her practical situation it does not follow that she is not obligated to Φ. In what follows, I will avoid using Darwall's talk of constitution. I think the idea is that certain motives are necessary, but not sufficient, conditions

---

[18] Darwall 1995, 15.

[19] Darwall 2002, 313-347. The relationship between obligations and ought claims will be discussed below. I have not specified the sense of ought in this passage since I am here only presenting a general account of empirical naturalist internalism, and am not yet saying how this view applies to Hobbes.

[20] There is a serious problem with Darwall's account that I cannot address here. Since obligation is *constituted* by motives that an agent *would have* if her beliefs about her practical situation were free from error, it seems Darwall must provide an account of what it is for something to be constituted by hypothetical motives.

for the application of the concept 'obligation'. Secondly, as noted, these hypothetical motives are not sufficient conditions. There also needs to be an explicit act of contracting, or laying down a right to Φ, in order for there to be a genuine (contractual) obligation.[21]

Darwall maintains that empirical naturalist internalism is a version of motive internalism. And while he never explicitly rules out the possibility of this being a version of reason internalism, there are a couple of things to notice in order to make sense of his claim. First, it may be that he had reservations about describing the view as a version of reason internalism because of some problems this characterization would encounter in his discussion of Hutcheson and, especially, Hume. There is, however, a more fundamental consideration and we can see this in his discussion of Hobbes. For Hobbes, it is crucial to specify that the necessary motives are *rational* motives. That is, they are motives that in fact respond to the dictates of right reason. Hobbes recognizes that the will can respond to reason in those creatures capable of rationality:

> And because in deliberation the appetites and aversions are raised by foresight of the good and evil consequences and sequels of the action whereof we deliberate, the good or evil effect thereof dependeth on the foresight of a long chain of consequences, of which very seldom any man is able to see to the end. But for so far as a man seeth, if the good in those consequences be greater than the evil, the whole chain is that which writers call *apparent* or *seeming good*. And contrarily, when the evil exceedeth the good, the whole is *apparent* or *seeming evil*…(L vi.57)

Elsewhere, Hobbes says, "men cannot put off this same *irrational appetite*, whereby they greedily prefer the present good" (DCv iii.32). He also explicitly denies that all appetites and aversions are rational when he defines 'emotions' in *De Homine*:

> Emotions or *perturbations* of the mind are species of appetite and aversion…They obstruct right reasoning in this, that they militate against the real good and in favor of the apparent and most immediate good, which turns out frequently to be evil when everything associated with it hath been considered. For though judgment originates from appetite out of a union of mind and body, it must proceed from reason. Therefore, although the real good must be sought in the long term, which is the job of reason, appetite seizeth upon a present good without foreseeing the greater evils that necessarily attach to it. (DH xii.1)

---

[21] That is, there needs to be an explicit act of contracting if it's true of her that she ought_O to Φ. I specify different senses of 'obligation' below.

Since it would be implausible to suppose that what is required for obligation is just *any* motive, desire or aversion, that an agent happens to have, I think it's more charitable to understand Hobbes as holding the view that the truth-maker for obligation claims is a particular kind of motive that an agent has (or would have), namely a rational motive. There is reason to think that Darwall would accept this claim. He says, "Although Hobbes defines 'obligation' independently of motive, it is critical to his view that obligation connect up with the agent's *rational will*, nonetheless."[22]

This suggests an obvious difference between internalism about goodness and internalism about ought judgments. Since internalism about goodness is grounded in whatever desires an agent happens to have, and these desires can, in principle, be for either immediate gains or longer-term goods, there is no reason to suppose that all of our motives are rational. We do not need to suppose that obligations will work in the same way if, as I am suggesting, the motives that are relevant are only rational motives. This suggests that the contemporary distinction between motive and reason internalism is not all that important for Hobbes. That is, insofar as Hobbes is committed to accepting a version of motive internalism (about ought judgments in general), he seems also to be accepting a version of reason internalism. On his account, there is a necessary relation between obligation, ought claims (in general), and an agent's motivational states. However, the motivational states to which they are connected are those such that an agent not only possesses an explanatory reason for her actions, but also possesses a justifying reason.

Now we have to consider Darwall's interpretation of Hobbes as holding a version of empirical naturalist internalism. Darwall claims that Hobbes, unlike the sentimentalists, emphasized an instrumental account of normativity that holds that "moral demands are binding just in case following them is necessary to accomplish ends that inescapably (if contingently) structure human deliberation – for Hobbes, self-preservation."[23] While Darwall is not clear on this point, this view develops out of Hobbes's account the nature of practical reason which is instrumentalist insofar as it maintains that

---

[22] Darwall 1995, 57; emphasis added.

[23] Ibid.

reason alone is not able to provide an agent with ends. The only practical function that reason has is to direct agents towards the means of satisfying their desires.[24]

There are a number of things that are unclear about Darwall's initial characterization of the Hobbesian brand of empirical naturalist internalism. First, as we'll see, Hobbes's definition of obligation in *Leviathan* does not explicitly appeal to an individual's motives and so it's not clear how to understand the claim that rational motives are a necessary condition for obligation. Secondly, Darwall refers to the end of self-preservation as "inescapable" but recognizes that such an end is nevertheless contingent, and it's not clear exactly what it is to have an inescapable end. Similarly, Darwall suggests that moral claims are binding when they are done in the service of *ends* that agents possess, but then only mentions the single end of self-preservation for Hobbes.[25] Does this imply that this is the only end that gives rise to obligation, or are there other ends that could do the same? In answering these latter two questions, we must also address how Darwall's responses could be consistent with Hobbes's claim that "though death is the greatest of all evils (especially when accompanied by torture), the pains of life can be so great that, unless their quick end is foreseen, they may lead men to number death among the goods" (DH xi.6).

Instrumental Reason View of Normativity

In order to address these questions, it will be helpful to begin with an explanation of Darwall's claim that Hobbes (unlike the sentimentalists) holds an instrumental reasons view of normativity. Darwall's view is an attempt to explain how it is possible for Hobbes to maintain that a law of nature is a "*dictate* of right reason" (DCv ii.1; emphasis added). This requires explanation because reason, on this account, is merely instrumental reason and so it's not clear in what sense reason is the kind of

---

[24] It is not clear to me whether Darwall seeks to identify Hobbes's instrumentalist theory of normativity with the instrumentalist account of practical rationality, or if he thinks the latter explains or justifies the former, or if there is some other relationship between them.

[25] It may seem as though Darwall is here giving up the hypothetical motives view of internalism. However, it's crucial for his interpretation to distinguish between the *ends* that individuals do in fact have, from the *motives* they would have given the correct use of their reasoning in the service of these ends.

thing that can issue dictates. The instrumental view of reasons holds that rational action is a matter of coherence with an individual's ends. That is, practical reasons for acting are given entirely by the ends that an individual already has, and reason has no other practical function that to direct agents in such a way as to achieve the satisfaction of her ends.[26] Darwall notes three principal reasons for attributing this view to Hobbes. The first two are textual, and the third appeals to broader implications from Hobbes's system.

First, Darwall cites the characterization of right reason that Hobbes offers in *De Cive*: "By right reason in the natural state of men, I understand not, as many do, an infallible faculty, but the act of reasoning, that is, the peculiar and true ratiocination of every man concerning those actions of his, which may either redound to the damage or benefit of his neighbours" (DCv ii.3). This passage clearly indicates that Hobbes denies the Thomistic view that the fundamental dictates of practical reason are self-evident.[27] However, it is less clear that Hobbes is here indicating a commitment to a theory of instrumental reason. This passage does not seem commit Hobbes to saying that the practical function of reason is limited entirely to means-ends judgments.

Darwall continues his case by appealing to the account of reason that Hobbes offers in Ch. v of *Leviathan* where Darwall quotes him as saying that reason is "nothing but reckoning." This is too quick. The passage from which Darwall draws this quotation says:

> Out of all of which we may define (that is to say determine) what that is which is meant by this word *reason*...For REASON, in this sense, is nothing but *reckoning* (that is, adding and subtracting) of the consequences of general names agreed upon for the *marking* and *signifying* of our thoughts; I say *marking* them when we reckon by ourselves, and *signifying,* when we demonstrate or approve our reckonings to other men. (L v.2)

The problem with interpreting this passage as indicating a commitment to a view of instrumental rationality can be seen in what Hobbes goes on to say about the use to which this reason is to be put: "The use and end of reason is not the finding of the sum and truth of one or a few consequences,

---

[26] I follow Gert in characterizing instrumental reasons in this way. See Gert 2001, 243.

[27] For a discussion on this point, see Schneewind 1998, 20-1.

remote from the first definitions and settled significations of names, but to begin at these, and proceed from one consequence to another" (L v.4). This is a description of the sort of reason that underlies Hobbes's account of science. "Science," says Hobbes, "is the knowledge of consequences" of the general names relating to the object of study. Demonstration is an important aspect of scientific method, which proceeds from the general names and their definitions, through syllogisms to "knowledge of all the consequences of names appertaining to the subject in hand" (L v.17).

Again, it is not clear that Hobbes is here committed to an instrumental view of rationality. He does continue by saying that the knowledge we get through science is such that "out of which we can presently do, we know how to do something else when we will, or the like, another time" (L v.17). However, while Hobbes does think that reason can be put in the service of ends in the manner required by the instrumental account, it is by no means obvious that he holds that it must be so employed. When Hobbes says that "reason, *in this sense*" is nothing other than the reckoning of consequences, he clearly indicates that there are other senses of reason besides this account of a demonstrative, inferential capacity.[28]

We still have to see the systemic considerations that lead Darwall to attribute the instrumental view of reasons to Hobbes. The basic idea is that, since Hobbes's metaphysics and philosophy of language require that all of our meaningful names refer in some way to something material, our meaningful propositions must refer in some way to material bodies and/or their properties. Given this, Darwall asks, "[h]ow…can the idea that reason dictates some action or other even have content?"[29] Or, as Gert puts the problem, "[h]ow can reason, if it simply means 'reckoning of the consequences of general names,' suggest convenient articles of peace [the laws of nature]? What general names or definitions of general names could plausibly have consequences that would lead men to agree on articles of peace?"[30]

---

[28] Both Gert, 1991 and 2001; and Rhodes 1992 and 2002 for an alternative account of Hobbes on reason.

[29] Darwall 1995, 58.

[30] Gert 2001, 249. As we'll see below, Gert argues that reason in the sense discussed here is insufficient to the task of generating these dictates.

Darwall thinks that Hobbes's answer is that reason is able to issue dictates only insofar as it operates in the service of ends that agents already have. That these ends are not themselves derived from reason is a fundamental tenet of the instrumental account. The "practical force" of reason derives from these ends: "It is in the transfer of motive force from ends to means by right reasoning that reason's dictates consist. We may call this the *instrumental reason* view of normativity."[31] This interpretation presupposes that the ends themselves have a "motive force" that can be transferred to means in the first place. If this is so, then the ends here must be construed as appetites or aversions, or in some way essentially related to appetites and aversions. This presents no problem on Hobbes's account since he maintains that everyone does have desired ends. Hobbes holds that our desires come to an end only in death: "Nor can a man any more live, whose desires are at an end, than he whose senses and imaginations are at a stand" (L xi.1). This seems to be the sense in which our desire for self-preservation is inescapable. More importantly, the rational dictates with which we are presently concerned do not recommend that individuals follow whatever course of action leads to some object that they desire. Rather, they recommend courses of action that lead to *the particular end of self-preservation*, as we can see in the definition of the laws of nature as dictates of right reason, "conversant about those things which are either to be done or omitted for the constant preservation of life and members, as much as in us lies" (DCv ii.1). Since Hobbes does often speak as if the avoidance of death is merely the object of a desire, there may be some reason for thinking that it is our desire for self-preservation that has the sort of foundational motive force required for Darwall's interpretation.[32]

---

[31] Darwall 1995, 58-9.

[32] It may be objected here that Darwall has not gone far enough to demonstrate the crucial aspect of the instrumental account of reasons, namely, that while reason does dictate means to achieving ends an agent has, reason does not dictate the ends themselves. Gert, for example, argues that Hobbes maintains that there is a sense of reason such that "it has an end of its own; avoidance of violent death" (1991, 13). The basic objection is that the mere fact that individuals do in fact desire their own self-preservation is insufficient to establish that reason has not, in some sense, given us this end.

Material Obligations and Divine Obligations

There appears to be at least three different senses of 'obligation' in Hobbes's moral and political theory.[33] All three of these are neatly summarized in the portion of *De Cive* where Hobbes considers our obligation to obey God:

> Now if God have the right of sovereignty from his power, it is manifest that the *obligation* to yielding him obedience lies on men by reason of their weakness. For that *obligation* which rises from contract…can have no place here; where the right of ruling, no covenant passing between, rises only from nature. But there are two species of *natural obligation*. One, when liberty is taken away by corporal impediments, according to which we say that heaven and earth, and all creatures, do obey the common laws of their creation. The other, when it is taken away by hope or fear, according to which the weaker, despairing of his own power to resist, cannot but yield to the stronger. From this last kind of obligation, that is to say, from fear or conscience of our own weakness in respect of divine power, it comes to pass that we are obliged to obey God in his natural kingdom; reason dictating to all, acknowledging the divine power and providence, *that there is no kicking against the pricks.* (DCv xv.7)

I have already provided a rough sketch of the sense of obligation that arises from contracts and will develop it in more detail in the next section. The other two senses of obligation are explicitly defined as versions of "natural obligation." Since Hobbes's aim here is to explain how it is that there could be obligations to obey God independently of any considerations about covenants, it is still possible for the obligations that are generated through covenants to count as versions of natural obligation. That is, the obligations that are generated through covenants might be one way in which one of the natural obligations is realized, namely by restricting liberty through hope or fear.[34]

So, we can see at least three distinct senses of obligation in Hobbes's claim, regardless of what the connection between them ultimately turns out to be. First, there is what I will call a material

---

[33] Cf. Oakeshott, Introduction to *Leviathan* 1962, lviii-lxi. Oakeshott claims that there are four kinds of obligation in Hobbes's theory. However, since Oakeshott follows Taylor in assigning a deontic sense of obligation to Hobbes, it's not entirely clear how well the distinction he finds in Hobbes maps onto my own.

[34] Warrender, for instance, maintains that all obligations reduce to natural obligations, so that the obligations that are generated through covenant are simply one way in which the natural obligation that is rooted in differential strength relationships gets realized.

obligation. Material obligations are states arising from physical impediments to actions.[35] These obligations will not be relevant since they do not apply to voluntary actions in virtue of the fact that an individual's conforming or not to physical laws, "the common laws of their creation," cannot directly be a matter of deliberation.[36] Secondly, there are what I will call divine obligations. Divine obligations are states arising from the difference in power between human beings and God. The idea here seems to be that we are obliged to obey God insofar as we fear the painful consequences of not following his decrees, and hope for the pleasant consequences by our obedience. Yet, Hobbes's language here is quite a bit stronger than this, for he says that the weaker "cannot but yield" to the stronger. It will be worthwhile to spell out in a bit more detail what Hobbes has in mind here.

It might be thought that Hobbes is attempting to analogize divine obligations to material obligations. That is, it may be thought that the weaker "cannot but yield" to the stronger in virtue of the fact that the stronger is physically constraining the weaker by literally binding, pushing, or otherwise prodding the weaker to do his bidding. On top of being unable to explain why Hobbes appeals to the notions of hope, fear, and conscience, this account does not provide an explanation for an individual's obligation to God. We are helped in our account by noticing that all obligations involve the taking away of liberty, and this is consistent with what Hobbes says elsewhere: "law and right differ as much as obligation and liberty, which in one and the same matter are inconsistent" (L xiv.3), and "where liberty ceaseth, there beginneth obligation" (DCv ii.10). However, if we appeal to the definition of liberty that we found in *Leviathan*, we end up with much the same view that attempts

---

[35] There is an interpretive problem about how to understand material obligations given that Hobbes seems to endorse the view that all obligations are the result of voluntary actions. For instance, he says "there being no obligation on any man which ariseth not from some act of his own" (L xxi.10). The same problem arises for divine obligation. I suggest that the charitable interpretation is to understand Hobbes as only intending to say that contractual obligations can only arise from voluntary actions. Some commentators have suggested that these non-voluntary senses of 'obligation' are "deviant." See Peacock 2010, 434 and Murphy 1994, 281

[36] Indirectly, individuals can deliberate about whether or not to construct material obligations for themselves. Odysseus deliberated about binding himself to the mast in order to resist the overwhelming temptation of the sirens, and so did indirectly deliberate about material obligation. Directly, however, once bound his being physically restricted from rushing off towards the sirens was not a matter of deliberation.

to draw a strong analogy between material and divine obligation. This is because the *Leviathan* characterization of liberty is the absence of external impediments, which suggests that the removal of liberty is the creation of external impediments. The characterization of liberty in *De Cive* is more helpful on this issue. The definition of liberty there is not much different from that provided in *Leviathan*: "liberty…is nothing else but *an absences of the lets and hindrances of motion*" (DCv ix.9). The crucial difference is the explicit discussion of what Hobbes calls "arbitrary" lets and hindrances "which do not absolutely hinder motion, but by accident, to wit, by our own choice; as he that is in a ship, is not so hindered but he may cast himself into the sea, if he will" (ibid.). Hobbes's meaning here is not entirely clear. It seems right that an individual is not normally physically restrained from throwing himself into the sea if this is his choice. In what sense, then, is there an arbitrary hindrance to action in this example?

We might imagine an impatient and frustrated passenger on a ship bemoaning, "I need to get to my business, but I am stuck on this ship!" A literal-minded, and understandably annoyed, friend might say, "You are free to leave the ship at any time." When the friend says that the impatient passenger is free, she means this in a very Hobbesian sense: there are no external impediments that keep him from casting himself into the sea. In the *De Cive* terminology, there are no absolute hindrances that keep him on board the ship. Nevertheless, there are arbitrary lets and hindrances that do keep him aboard: he would reasonably respond to the friend by saying that, even though he is *strictly speaking*, free to leave, there is another sense in which he is not free to leave. When we evaluate the freedom of our actions according to this looser standard, we incorporate into our judgments the *desirability* of the available alternatives. Our actions are "hindered," not absolutely, but by our "choices" insofar as the options we have open to us are not equally desirable. In this case, it's not merely that the man does not presently care to go for a swim, but that his more fundamental desire (to attend to his business) would be frustrated if he were to cast himself into the sea. He is still at liberty to do so, but there is a sense in which his liberty is restricted since the alternative that is open to him would not allow him to pursue other ends.

This is a rather idiosyncratic understanding of the notion of a *hindrance* to actions. Nevertheless, such a notion is crucial for understanding Hobbes's views of obligation. He says that hope, fear, and despair are all part of the taking away of liberty in divine obligation. Hope is an appetite with an opinion of attaining the object of our desire, while despair is an appetite without an opinion of attaining (L vi.14-15).[37] Fear is an aversion to something with "an opinion of hurt from the object" (L vi.16). When Hobbes says that liberty can be taken away by hope or fear, then, he is claiming that particular opinions regarding the likely consequences of certain courses of action are connected up with certain appetites and aversions in such a way that the individuals whose opinions these are will choose to act in a particular way. We can also say that their liberty is hindered in these cases insofar as their range of choices is severely restricted: the power of God is irresistible. For instance, if we suppose that God is the sort of being that is capable of imposing eternal punishment on those who violate his commands, and some individual knows that God is this sort of being, then, in virtue of her aversion to punishment and her opinion that such punishment will likely come about as a result of her violating God's commands, she will choose to follow God's commands. Her liberty is restricted by the irresistible power of God insofar as violating his commands does not represent a viable option for her.[38] Note, too, that fear is not the only passion that can constrain an individual's will in this way. Hope will work in a similar fashion, the difference being that an individual will desire the eternal rewards of following God's commands, and will have an opinion that following God's commands will likely bring about such rewards.

---

[37] The phrase "without opinion of attaining" is ambiguous between "having an opinion that attaining is unlikely" and "having no opinion of attaining whatsoever." Presumably, Hobbes means to characterize despair as an appetite with an opinion that an attempt to attain the desired object will likely be unsuccessful. If an individual had no opinion about attaining the desired object, then Hobbes would have no way of distinguishing despair from appetite *simpliciter*.

[38] It might be thought this suggests that the atheist does not have any divine obligations. Since the atheist will not fear God's punishments, the atheist will not reason along these lines. This might be thought to be inconsistent with Hobbes's view, since he seems to hold that atheists are obligated to follow the laws of nature. See Rhodes 2002, 48. I maintain that, while Hobbes could have denied that atheists have divine obligations, this does not entail that he would have also denied that atheists have moral obligations to obey the laws of nature. This is because I deny that the obligation to obey the laws of nature derives from God's commands. Atheists are bound to follow the laws of nature regardless of their beliefs about God.

This account can only be plausible if we suppose that individuals can have good reason for thinking that their opinions about eternal rewards and sanctions are likely to come about as a result of particular courses of action. How Hobbes can justify this claim is beyond the scope of this discussion. I am here only claiming that if God exists, then this is how we should understand divine obligation. I discuss divine obligations only to see whether or not the account Hobbes gives of them is similar to his account of contractual obligations, and, if so, how these similarities relate to the account of internalism. The first thing to notice about divine obligations is that Hobbes is quite insistent that they arise out of the difference in power between God and humans. He says, "the weaker, despairing of his own power to resist, cannot but yield to the stronger." Despair, again, is an appetite accompanied by the opinion that an attempt to attain the object desired will likely be unsuccessful. It is the power of God that makes it reasonable for an individual to think that she could not escape, say, the sanctions that would follow violating God's commands. So, for Hobbes, divine obligation is just the restricting of one's actions though the will because one sees as inescapable the consequences of divinely secured rewards and punishments. Since the will consists, partly, in appetites and aversions, and since appetites and aversions are motives, it follows that divine obligation is necessarily related to motives. The power of God structures an individual's deliberations in such a way that any *reasonable* individual will be moved to follow God's commands.

Contractual Obligations, 'Ought' Claims, and Internalism

What sort of necessary connection obtains between an individual's judgments regarding particular courses of action and her motivational states in the case of divine obligations need not concern us here. Again, the aim here is to see if understanding divine obligations can help us make sense of contractual obligations and their relationship to rational motives. In order to see the connection, it will be helpful to specify the relationship between contractual obligations and the various senses of 'ought' discussed above. Contractual obligation is defined in terms of the abandonment or the transference of a right, and Hobbes says of an individual who lays down a right in either manner "that he *ought*, and it is his DUTY, not to make void that voluntary act of his own, and

that such hindrance is Injustice, and Injury, as being *sine jure*, the right being before renounced or transferred" (L xiv.7). It follows analytically from the definition of contractual obligation that one ought, in the ought_O sense, to keep covenants made. It does not follow from the definition, however, that one ought, in the ought_L sense, to keep covenants. This is why Hobbes goes on to offer the third law of nature as a rational dictate that one ought_L to keep covenants made.

In the case of divine obligation, what we see is that individuals ought_L to obey God. This is because there can be no contractual obligation to obey God: "[t]o make covenant with God is impossible" (L xiv.23). We saw above how this connected with an individual's motivational states. If Hobbes thinks there is a strong connection between divine and contractual obligations, then he has to have some way of connecting the ought_L claims with the ought_O claims that follow analytically from laying down a right. According to Darwall, the connection will be secured through the motives that are necessarily entailed by the truth of the ought_L claims. As he understands the connection, "joint undertakings are valid contracts, and hence obligating, if, and only if, performance is dictated by the law of nature. An obligation 'ought' will exist only if a law 'ought' does."[39] Now, a law 'ought' applies to an agent only insofar as the action it prescribes has as its end the self-preservation of the agent (or has as its end an object of the agent's desire). The connection here can be better specified by considering again why it is that individuals set aside their original right of nature. This is done because it is the only way for individuals to escape from the natural condition where the end of self-preservation is overwhelmingly likely to be frustrated. I noted above that setting aside a right involves divesting oneself of the liberty to do *whatever*, in one's own judgment, is required for self-preservation. If we understand this divesting of liberty in terms of removing ourselves as external impediments, then Hobbes sounds like he is grounding contractual obligation in a material obligation. This, however, cannot be the case since, again, material obligations cannot directly be the subject of deliberation.[40] So, the divesting of liberty here should be understood in the same way that it was

---

[39] Darwall 1995, 71.

[40] It might be thought that the material sense of obligation is important for thinking about the role of the sovereign in creating obligations. The thought is that the sovereign's responsibility is to enforce

understood in the case of divine obligation. That is, an individual divests herself of some liberty when she wills, by reasoning about the inescapable consequences of her actions, a particular course of action. Her obligation to perform just is her will to perform insofar as she does something (contracts) that restricts the rational courses of action that are open to her. I will have more to say about why contracting can be thought of as restricting the rational courses of action that are available to an individual in my interpretation of the fool below. For now we can see that Hobbes is an internalist insofar as he holds that an agent can be *contractually* obligated only when she is in certain motivational states, namely those states that accompany her true beliefs regarding the necessary means to her self-preservation. If she were to use signs expressive of her will to contract, and were her beliefs about the means to preservation free from error, she would have some motive to keep her agreement.

If this interpretation is correct, it offers an explanation for the crucial role that the sovereign plays in Hobbes's system. The problem, as noted previously, is that there is no security in the natural condition, so that the performance of contractual agreements will likely make an individual in such a condition the prey of others:

> But because covenants of mutual trust where there is a fear of not performance on either part….are invalid, though the original of justice be the making of covenants, yet injustice actually there be none till the cause of such fear be taken away, which, while men are in the natural condition of war, cannot be done. Therefore, before the names of just and unjust can have place, there must be some coercive power to compel men equally to the performance of their covenants, by the terror of some punishment greater than the benefit they expect by the breach of their covenant… and such power there is not before the erection of a commonwealth. (L xv.3)

The role of the sovereign, then, is to compel individuals to perform their covenants made and to do this through fear. The sovereign must make people believe that it is likely that violation will yield

---

covenants and to do so by creating physical restrictions to one's actions (imprisonment, death, etc.). I do not think that this interpretation is entirely successful. Obligations involve the restriction of one's liberty to do or not do some act. The physical restrictions on one's liberty that follow from violation, however, occur after the fact. The physical punishments imposed by the sovereign for violating his commands, then, come into the account after one has already failed to uphold one's obligation (understood as a condition involving the restriction of liberty). The same is true of divine obligations. So, neither divine obligations nor contractual obligations should be thought of as kinds of material obligations.

punishment, an end to which they are averse. In other words, the sovereign must, with respect to contractual obligations, act like God acts with respect to divine obligations. Indeed, Hobbes does sometimes speak of the sovereign in these terms: "This done, the multitude so united in one person is called a COMMONWEALTH…This is the generation of that great LEVIATHAN, or rather (to speak more reverently) of that *Mortal God* to which we owe, under the *Immortal God*, our peace and defence" (L xvii.13). The closer the analogy between the sovereign power and God, the stronger an individual's judgments that punishment inevitably follows violation will be.

Assuming a very close analogy between the sovereign and God, we still need to specify how exactly the truth of $ought_L$ claims connects up with an agent's rational motives. The connection is secured through the instrumental reasons view of normativity. This view holds, again, that the normative force of practical ought claims, in this case $ought_L$ claims, derives from the motivational force of ends that agents already have. Given that agents have as an end their self-preservation, the rational judgments that certain means are necessary to secure this end will thereby inherit the motivational force of the end itself. Now, the natural condition is one where reasonable agents will judge that the end of self-preservation is very likely to be frustrated. Since the establishment of a commonwealth through covenants with each other is the necessary means to escape the natural condition, rational agents will judge that such a means $ought_L$ to be pursued. Since the laws of nature specify the way to establish a commonwealth, rational agents will judge that they $ought_L$ to follow the laws of nature. However, these judgments, especially those concerned with the keeping of covenants, will only be rational when there is a sovereign in place to make violation more undesirable than non-violation. We are assuming here that this is the case. So, the motivational force of the end of self-preservation is transferred to the keeping of covenants through an agent's rational judgments that keeping such covenants is necessary to achieve that end. That is, they $ought_L$ to keep their covenants insofar as they have some motive to do so.

There is another way of seeing the connection between $ought_L$ claims and motivational states. $Ought_L$ claims will always be connected to what is good for an agent. This is because, as we've seen, $ought_L$ claims, if they are to be the sorts of things that can *dictate* courses of action to agents, depend

on the ends that agents have. Insofar as these ends are desired, they will be the sorts of things that agents will call good. Even though Hobbes recognizes that good and evil are not objective properties of objects, and can therefore vary greatly between different people depending on their constitution, their habits, their environment, etc., nevertheless, the inescapable end of self-preservation is, in most cases, counted among the goods. According to the instrumental reasons view of normativity, if an end is desired, then the necessary means to that end is desired as well, insofar as the agent is rational. This, then, is how Hobbes can claim that "all men agree on this, that peace is good; and therefore also the way or means of peace…are good" (L xv.40). Given internalism about good, then, it follows that the truth of an $ought_L$ claim depends on motives an individual has in virtue of her right reasoning.

Internalism and Hobbes's Response to the Fool

According to Darwall, it is not the case that one $ought_O$ to do something if it's not also the case that she $ought_L$ to do it. This amounts to the claim that contractual obligations are only valid when they $ought_L$ to be done. The above account of the relationship between contractual obligations and what an individual $ought_L$ to do depended on a very close analogy between the sovereign power and God. Namely, the sovereign is responsible for making it rational for individuals to believe that certain undesirable consequences will follow from the violation of covenants. If the account is to be successful, then we should be able to say about the individual contemplating violation that he despairs of his power to resist. That is, we should be able to think of potential violators as having a desire to resist but nevertheless having an opinion that such resistance will not be successful. Where such an opinion is lacking, we would have to conclude, as we would for potential violators of God's commands, that they *irrationally* judge that some good can come about by kicking against the pricks.

This is the situation that the fool presents us with, since the fool thinks that it can be rational to act against the sovereign's decrees by breaking valid covenants.[41] One way of thinking about the

[41] Of course, by violating a covenant, the fool is indirectly acting against the sovereign. Since no individual contracts directly with the sovereign, no violation of covenant is a violation of a covenant with the sovereign. It is, however, an indirect violation insofar as the sovereign is tasked both with enforcing covenants as well as creating civil law. The natural law, according to Hobbes, gets encoded into the

fool's objection, then, is to see it as a response to the close analogy between the sovereign and God. Basically, the fool denies that the sovereign has the sort of irresistible power that God has. While the sovereign may be able to enforce a vast majority of covenants, and may be able to impose rather severe punishments for their violation, it is implausible to suggest that the sovereign is either as omniscient or as omnipotent as God (or even nearly so) on this picture is supposed to be. As a consequence, the fool thinks that he may, with reason, violate a covenant in a commonwealth because he thinks that the benefits of gaining are sufficiently great so as to outweigh the consequences of being caught, and the fool also recognizes that the probability of being caught may be sufficiently low to render the bad consequences moot. So, the fool maintains that he is under neither divine nor contractual obligation, since he denies that God exists and maintains that contracts can reasonably be violated when the payoff, weighed against the risks, for doing so is sufficiently great.

We have already seen some of the main interpretive strategies that contemporary commentators employ with respect to the fool's objection and Hobbes's response to it. What's left to be seen is how these strategies can connect up with the account of internalism to be found in Hobbes. In what follows, I present what I take to be Darwall's argument for adopting something like Kavka's interpretation.[42] I provide some additional considerations to this argument that help to explain the fool's practical situation. The main thrust of the argument is that, since Hobbes's internalism posits a necessary connection between contractual obligation and the end of self-preservation, and given the unreliability of individual prudential calculation, what agents have most reason to do in order to secure preservation is to keep covenants made.

For our purposes, we can focus our attention only on what Hoekstra calls the Silent Foole. This is the Foole that deceives without detection and who makes no public declarations that violating

---

sovereign's civil law so that the sovereign does command that covenants be upheld (and has the power to enforce this command).

[42] I do not mean to be taken to endorse Kavka's interpretation. I mentioned one fairly serious worry with the view above. It is not clear that internalism is a decisive issue in choosing between alternative interpretations of Hobbes's exchange with the fool.

covenants may be within reason. We should restrict our attention to the Silent Foole since both Hampton and Hoekstra think that Hobbes is in agreement with him. Kavka and Darwall, however, deny that this is so. Kavka maintains that, even if an individual has good epistemic reason to believe that violating will be beneficial to her, it is nevertheless more reasonable overall for her to keep her covenants. Since rationality is not spelled out in terms of case-by-case, expected utility maximization, Kavka can recognize that an instance of violating may bring about greater expected utility, while nevertheless failing to be the rational thing to do. Darwall notes that it is not the case that Hobbes believes that an individual should do whatever it is she has most reason to believe will lead to her greatest expected benefit, or what will yield the highest likelihood of her self-preservation, given the available evidence. Hobbes is therefore not committed to the view that violating a covenant can never be what an individual reasonably believes, given the available evidence, will bring about the highest expected utility. The main point that Hobbes is emphasizing is that what we have evidence to reasonably believe may nevertheless be false: "It can happen, therefore, that although an agent reasonably believes on evidence that violation will pay, this is disastrously mistaken."[43]

Since the stakes are incredibly high in cases of covenant violation, Hobbes does not think that the standard of rationality should be one based on the reasonability of belief relative to available evidence. The costs of being found out for violating are so high that "right reasoning in pursuit of self-preservation does not recommend violating covenant even when what it would be most reasonable for the agent to believe is that violation would benefit (even, indeed, when she takes account of her own fallibility)." [44] This, as both Kavka and Darwall note, does not mean that Hobbes would disagree with the fool concerning violations under conditions of certainty. It would be contrary to reason to *not* violate when we *know* that such violation will be best for us. However, Hobbes insists that we never are in such a position to *know* this. What's more, given the incredibly high stakes, Darwall suggests that even relying on estimations of the probability of certain courses of actions is not a practically

---

[43] Darwall 1995, 76.

[44] Ibid.

rational decision procedure. That is, decisions regarding potential violations should be thought of as decisions under conditions of uncertainty, rather than as decisions under conditions of risk. This answers part of the objection that I raised against Kavka in the previous chapter. We have pragmatic reason to judge that decisions regarding potential violations are decisions under uncertainty, and so the relevant probabilities, while perhaps epistemically available, are not pragmatically available (so to speak). Valid covenants create states of affairs that change the rationality of following one's own individual judgment regarding what it is best to do in each particular case. What it becomes rational to do with respect to covenants is to commit to keeping them, and to treat covenants *as if* they created, as Darwall says, "categorical constraints."

There are two parts to Hobbes's explanation of the fool's *obligation* to keep covenant on this view. Both appeal to the characterization of ought_O and ought_L claims discussed above. When an individual contracts to do something, it is true of her, as a matter of definition, that she ought_O to do it. Also, it is true of her that she has laid down, or otherwise transferred, her right to something. In particular, her right of nature is qualified or restricted as soon as she contracts. The nature of this restriction is two-fold. First, and most apparently, by contracting she no longer has the original right of nature that entails she has a right to all things: there is now, after contract, something to which she does not have a right. Again, this right is to be understood in terms of her liberty to do what she judges to be the aptest means to her preservation. Moreover, right reason in the state of nature, prior to any contracts, dictates that she rely on such judgments since no alternative is available. This points to the second way in which the right of nature is qualified in contracts: where contracts are involved right reason does not dictate that individuals should rely on their own judgment about the aptest means to their preservation. In the case of contracts, says Darwall, "right reasoning recommends keeping contract regardless of what the agent believes the relation to be between doing so and his own preservation or good."[45] What this means is that, when an individual ought_O to do something, she rationally should no longer rely on the deliverances of her own judgments about whether or not

[45] Ibid., 78.

keeping covenant will *in this case* be in her self-interest. By contracting she has performed an action whereby her own judgments about this particular case simply become irrelevant to the rationality of her action.

This is the first part of Hobbes's response to the worry about the fool's motivations. The second part of the response depends on the further claim that, according to Hobbes's internalism, it's true that someone $ought_O$ to do something only if it's true that she $ought_L$ to do it. More importantly, it's true that she $ought_L$ to do something if and only if it's true that "she would be moved to do it if her empirical beliefs about her practical situation were error free." So, the question arises if the fool would be moved to keep covenant if his beliefs about his practical situation were error free? I think the answer is "yes," and I think we can see that this is Hobbes's position by pointing again to the two-fold restriction of right that comes along with covenanting. First, by contracting one is laying down or otherwise transferring a right to some thing or to do some act. The act of *making* the contract, as opposed to the act of keeping it, is still to be determined by the standard of reasonability relative to available evidence. An act of laying down or transferring a right to some thing, or to do some act, is reasonable when doing so is judged by available evidence to yield the highest expected benefit compared to alternative actions. In such cases we will say that it's true that the individual $ought_L$ to make the contract. If it's true that she $ought_L$ to make the contract, then it's true that she is moved to make the contract in virtue of the role that contracting plays in her self-preservation.

This part of the reply seems to be fairly plausible: the fool has covenanted and so must, by use of his own judgment, think that doing so was what conduced most to his ends, and is thereby provided with motive to do so. However, the fool may have covenanted disingenuously, without ever having formed an intention to follow through. In this case, it would be strange to think of the fool as being moved to lay down his right to some thing, or to do some act. What the fool is moved to do is to make it seem as if he has made an agreement with the intention of following through. The most that we can say of the fool on this account seems to be that he had motive to make (or seem to make) a covenant, but no motive to keep it. If this is so, then we have failed to address the objection.

If the second part of the reply is to be successful, it must show how it could be that even a disingenuous fool could be moved to perform a covenant that he had no intention to keep. This is indeed a tall order. The reply again appeals to the second way in which contracting restricts our original right of nature, viz. by changing the conditions of choice in such a way that right reasoning no longer recommends relying on one's own judgment regarding the rationality of keeping or violating a covenant. The fool's practical decision to keep or violate a covenant is, trivially, conditional on his having made a covenant. This, however, is anything but trivial when it comes to the conditions under which such a choice is to be made. Covenants are powerful tools. They are the principal means by which individuals are capable of escaping the natural condition, the principal tools whereby mutually beneficial social interactions are ensured, and the principal tools for keeping a society from deteriorating back into the state of nature. Their power does not derive from any non-natural normative force, but lies solely in their ability to establish the trust and security that people require in order to engage in beneficial, cooperative activities. Hobbes nowhere suggests that the intention to keep a covenant is a necessary condition for making a valid covenant.[46] Once the fool has covenanted, he has *in fact* changed the conditions underlying the rationality of his choice. He is using one of the most powerful tools that rational agents have at their disposal, and this is so even if he has every intention to misuse such a device.

This account does not imply that every apparent covenant is an actual, valid covenant. There are still conditions under which covenants can be made void. The account is instead insisting that, where someone makes every effort to make their covenant appear to be genuine, the mere fact that they have no intention to keep it does not itself make their covenant void. Other cases of promise making seem to be like this. If I intend to deceive a host about my attendance to his party by sending in my RSVP with the box marked 'yes', knowing full well

---

[46] This is distinct from that claim that a *sign* that the other party has no intention to keep covenant is sufficient to make the covenant void. The covenant is made void not in virtue of the other parties lacking an intention to follow through, but in virtue of the fact that the first party reasonably believes that such an intention is absent.

that I will not go to the party, this fact alone does not render my RSVP invalid. Unless my host has some reasons to suspect my deception, my RSVP will function no differently in his deliberations than will the RSVP of his most honest guests.

The fact that the fool really is covenanting is significant because it implies that he really has changed his decision situation in such a way that his reliance on his own judgment about the rationality of keeping or violating is no longer the recommended decision procedure. This is one of the features of his practical situation and it is one about which he is quite mistaken. If the fool were to come to see that his use of covenant brings along with it a restriction of his right such that rationality is no longer to be determined by his own judgment, then, Hobbes suggests, the fool would see that treating covenants as if they created categorical constraints would lead to what's most in his interest. That is, right reasoning recommends the keeping of covenants even when one judges that violation will benefit. So, it's true that one who contracts ought_L to keep his agreements. Given that ought_L claims are true only if the agent would have some motive to comply with them, Hobbes seems to be committed to the view that the fool would be moved to keep his agreements if his beliefs about his decision situation were free from error. The foregoing discussion suggests that the main error that the fool makes is in his insistence that the rationality of *keeping or violating* covenants is no different from the rationality of *making or not making* covenants in the first place.

If the fool were to see just how powerful covenants are, he would not treat them as a means of self-interested deception. The truly enlightened fool would simply give up on the attempt to use contracts as a means of deception: they are not effective tools for that job. If the fool instead came to see exactly how contracts change his practical situation, he would thereby acquire some motive to keep his contracts. Hence, the fool is obligated to keep his covenants. He is obligated in part because, were he to see how it is that contracts work in rational deliberation, he would have some motive. This is why it's important to interpret Hobbes's internalism as a view about hypothetical or dispositional motives, rather than actual motives. So, we can see that Hobbes's internalism is fully consistent with his response to the fool. I have not argued that Hobbes's response to the fool is successful, nor have I

argued that Hobbes's brand of internalism is plausible. My aim has been to show what Hobbes is up to in his response to the fool and to clarify how this response fits in with his internalist commitments.

A Hobbesian Response to the Amoralist's Challenge

As we saw in the introduction, Brink argues that there is an inconsistency between adopting certain versions of internalism and holding the question "why be moral" to require a substantive response. His claim is spelled out in terms of the possibility of an amoralist or in terms of the intelligibility of the amoralist's demands. The amoralist can be thought of as "someone who recognizes the existence of moral considerations and remains unmoved."[47] If one accepts the view that recognizing moral considerations brings along with it some motive, then it seems as if there is at least a *prima facie* tension between the possibility of the amoralist and certain versions of internalism.

Taking the fool as Hobbes's representative amoralist, we are in a position to see how Hobbes could respond to Brink's argument here. First, Hobbes would present his answer to the question "why be moral?" He will understand this question as asking what reason we have to act virtuously, where acting virtuously is understood in terms of following the laws of nature: "For moral philosophy is nothing else but the science of what is good and evil in the conversation and society of mankind… Now the science of virtue and vice is moral philosophy; and therefore the true doctrine of the laws of nature is the true moral philosophy" (L xv.40). Taking the virtue of justice, defined by the third law of nature, as our guide, Hobbes would say that we ought ($ought_L$) to be just because doing so is what is most in our interest. It is important to stress here that being just is what is most in our interest insofar as we are considering ourselves *in relation to others*. I cannot ignore other people when I think about the rationality of moral action because, by definition, moral action concerns "the conversation and society of mankind." So, we $ought_L$ to be just because doing so is what is most in our interest as one individual among many others.[48]

---

[47] Brink 1989, 46.

[48] I emphasize this, seemingly trivial, point because it is by recognizing my relation to others that contracts, the chief mechanism of justice, can be seen to alter the decision situation as suggested above. Covenanting is my way of establishing myself as a trustworthy, or not, collaborator. Having trust is vital

Hobbes would then go on to reaffirm his commitment to internalism. He maintains that an individual is under an obligation to Φ in virtue of the fact that, other things being equal, he would be moved to Φ were his practical beliefs free from error. That is, he $ought_O$ to keep his obligations iff he $ought_L$ to do so. The fool, as our representative amoralist, seems to suggest that he recognizes his obligations ("He does not therein deny that there be covenants, and they are sometimes broken, sometimes kept, and that such breach of them may be called injustice, and the observance of them justice"), but he nevertheless remains unmoved to perform them, or at least questions the reasons he has for keeping them ("he questioneth whether injustice…may not sometimes stand with that reason which dictateth to every man his own good"). The fool is thereby attempting to drive a wedge between $ought_O$ and $ought_L$ claims, and this is precisely what we would expect to follow from a denial of Hobbes's internalism.

I think the best way to understand how Hobbes would reply to Brink's concern here is to see how Hobbes can deny that the fool really does *recognize* his moral obligations. The fool, having covenanted, most certainly does recognize that he is obligated to perform, since this obligation follows analytically from the act of covenanting.[49] However, what the fool fails to see is that he is under a genuine moral obligation to follow through. Again, the fact that it is a moral obligation depends, at least in part, on the fool's relationship to other people. In particular, since contracting is the most valuable tool that individuals have at their disposal for establishing the security and trust they need in order both to escape from the natural condition and to prevent their return to it, the sorts of relationships that the fool needs to be concerned with are among the most important that human beings can have with each other.

---

for my self-preservation. Without focusing on the role that other people play in my deliberations, it's hard to see how there could be any reason to be just.

[49] Hobbes's response to the fool would indeed be much less controversial, and a great deal less interesting, if it were only about the analytic inferences that can be made from the fool's having covenanted.

As we saw above, the fool mistakenly believes that, after covenanting, he can still use his own judgment as the standard of rationality. This is, however, to ignore the fact that part of covenanting involves giving this up as the standard of rationality. Covenanting changes the decision situation in such a way that mutually cooperative actions become possible because it becomes rational for individuals to follow general rules. Given that all of this is tied up in the notion of contracting, and depends on the relationship between $ought_O$ and $ought_L$ claims, and given that the fool fails to recognize these relationships, there seems to be a genuine sense in which the fool fails to recognize his moral obligations. So, Hobbes's version of the amoralist is, in fact, not really an amoralist. He is, instead, someone who has not fully grasped the nature of moral obligations, and has instead blindly made covenants without recognizing the significance of what he's done.

CHAPTER 3

HUTCHESON: MORAL SENSE AND PASSIONS

Introduction

It was standard practice in post-Hobbes British moral philosophy to attempt to provide a complete refutation of Hobbes's view. This tendency is exemplify in Francis Hutcheson's early work where, as I will argue below, the development of his moral sense theory arises straightforwardly from his rejection of, what he took to be, Hobbes's view. Nevertheless, Hutcheson, like Hobbes before him and Hume after him, had fairly strong empiricist commitments. However, Hutcheson is explicit about some of his theological commitments and there doesn't seem to be any dispute, unlike with Hobbes, whether these commitments are genuine. This does not undermine his naturalism. A moral theory is a naturalistic, in the sense that is relevant here, if it attempts to "account for normativity, in fully natural terms, without relying on any supernatural posits."[1] Even though God does play a role in Hutcheson's moral theory, it is not the case that Hutcheson attempts to understand moral obligation in terms of God's commands. God enters into the explanation of why we have the psychological faculties, including the moral sense, that we do, and it is these faculties that figure into the account morality and obligation. So, Hutcheson is an empirical naturalist, and as we will see his brand of empirical naturalism is thoroughly anti-Hobbesian.

In this chapter, I will lay out some of Hutcheson's fundamental views. In the first section, I briefly describe Hutcheson's empiricist commitments and show how they figure into his important division of the senses. I then present Hutcheson's account of goodness, both moral and natural, and its relation to his rejection of egoism. I also have a brief discussion of an interpretive controversy regarding the question of whether Hutcheson is best understood as offering a subjectivist, emotivist, or realist theory. I then turn to Hutcheson's rejection of rationalism and

---

[1] Darwall 1995, 14-5.

its important appeal to the distinction between exciting and justifying reasons. I conclude with a discussion of the various elements of Hutcheson's moral psychology in order to get a clear picture of the role that reason plays in the formation of calm affections. I will take up the question of moral motivation, the relationship between motivation and rational action, and the "why be moral" question in the next chapter.

Hutcheson's Empiricism

To begin, many of Hutcheson's argumentative strategies take for granted his epistemological commitments. Hutcheson was, broadly speaking, a Lockean.[2] His Lockean commitments led him to accept the view that all simple ideas are traceable to the operations of either sensation or reflection. While Hutcheson does refine Locke's account of reflection, or inner sensation, his modifications are not large departures from Locke's central commitments (e.g. he eschews all appeals to innate ideas or principles). Also, Hutcheson's Lockean commitments lead him to endorse a particular conception of philosophical method, namely one which emphasizes introspection. On this second point, Hutcheson says, "in this inquiry [regarding the operation of the affections and the passions] we need little reasoning or argument, since certainty is only attainable by distinct attention to what we are conscious happens in our minds" (EPA I.i). Hutcheson's point here is that there are some contents of our minds, specifically affections and passions, about which we can have certainty when we pay them "distinct Attention." As we will see, the affections and passions play an essential role in

---

[2] I do not mean by this that Hutcheson uncritically accepted wholesale Locke's theory. Such a suggestion would ignore the places where Hutcheson both explicitly and implicitly criticizes Locke's view. A number of commentators have denied that there is a significant sense in which Hutcheson was a Lockean. Scott (1900, 19), and to lesser degrees Norton (1985, 398-405) and Mautner (1993, 46) all advocate some version of the view that Hutcheson's commentators do him a disservice by reading in more of Locke than Hutcheson would likely accept. In what follows, nothing will depend on whether or not Hutcheson accepted many of the details of Locke's account. That is, none of the arguments in this chapter will appeal to the views of Locke as providing some reason in favor of a particular interpretation of Hutcheson. I mention the Lockeanism here only for the sake of illustrating the kinds of epistemological and methodological constraints Hutcheson adopted for his theory.

Hutcheson's theory, and so it is understandable why Hutcheson relies as he does on the role of introspection in moral philosophy.[3]

As a Lockean, Hutcheson is committed to the view that the source of all of our simple ideas is either sensation or reflection, though the conception of these powers undergoes some revision in Hutcheson's hands. Hutcheson maintains that, "there seems to be some Sense or other suited to every sort of Objects which occurs to us, by which we receive either Pleasure, or Pain from a great part of them, as well as some Image, or Apprehension of them: Nay, sometimes our only Idea is a Perception of Pleasure, or Pain" (EPA I.i). The senses are, in part, capacities for feeling distinct kinds of pleasure or pain, and are only sometimes the occasions of images, tactile sensations, smells, and etc. Hutcheson's general definition of a sense is a "Determination of our Minds to receive Ideas independently on our Will, and to have Perceptions of Pleasure and Pain" (ibid. 17, R 356).[4]

Hutcheson distinguishes five broad classes of senses according to the kinds of objects that are, partly, causally responsible for giving rise to pleasures and pains. He later goes on to specify distinct original, by which he means intrinsic, desires and aversions that correspond to each set of pleasures and pains:

1. External sense: These are the familiar senses of sight, hearing, taste, etc., but Hutcheson

---

[3] Neither Locke nor Hutcheson were particularly worried about the skepticism that was lurking behind the views they accepted. Locke says, "The notice we have by our senses, of the existing of things without us, though it be not altogether so certain, as our intuitive Knowledge … yet it is an assurance that deserves the name Knowledge. If we persuade our selves, that our Faculties act and inform us right, concerning the existence of those Objects that affect them, it cannot pass for an ill-grounded confidence: For I think no body can, in earnest, be so sceptical, as to be uncertain of the Existence of those Things which he sees and feels" (ECHU IV.XI.3). See also, Lowe 2005, 54.

[4] Hutcheson maintains this definition of sense throughout his early writings. See *Inquiry* Preface, p.xiii; I.i. Raphael argues that this conception of a sense is too broad which in turn explains why Hutcheson ends up with an overly permissive account whereby "every faculty except the will counts as a sense" (Raphael 1947, 20). Jensen too remarks that Hutcheson's introspective methodology leads to his "surprising discoveries of a new sense at the rate of about one every four years!" (Jensen 1971, 40). Both Raphael and Jensen seem to overlook that Hutcheson is merely describing psychological dispositions of normal human beings. Why should we expect there to be an in principle upper limit on the number of dispositions to which human beings are susceptible?

claims that there are seven to ten of them, not five. Hutcheson says of these that they are "universally known" (EPA I.i, R 356). He maintains that our first ideas of pleasure and pain are acquired through the external senses, and identifies what he calls "natural goodness" and "natural evil" with these perceptions of pleasure and pain respectively (*Inquiry* intro., R 304).[5] Hutcheson upholds Locke's primary/secondary quality distinction, but claims that only the secondary qualities are properly the objects of external senses.[6] The external senses give rise to desires for sensual pleasures, and aversions to "the opposite Pains" (EPA I.i).

2. Internal sense: The internal sense is that sense by which we take pleasure in uniformity and regularity, as well as pleasures arising from "Grandeur and Novelty" (EPA I.i, R 356). The internal sense is responsible for our perception of beauty.[7] It is important to note that, while similar in their operation, Hutcheson conceives of the moral sense as distinct from the internal sense *per se*. There are, indeed, some places where Hutcheson seems to suggest that all those senses that are not external are internal senses, as in the following:

---

[5] This has led some commentators to maintain the position that Hutcheson was a hedonist. According to Jensen, Hutcheson does not use consistent language when talking about this point and so sometimes suggests, as he does here, that natural goodness is to be identified with pleasure as a mental state. Other places what seems to be naturally good are those objects, events, relations, etc. which give rise in us to various pleasures. According to Jensen, this latter view is what is suggested "at all of the crucial places" (Jesen 1971, 21). All of the crucial places, that is, besides the purported definition of natural goodness. In either case, I do not need to settle on an interpretation of Hutcheson's theory of intrinsic value in what follows and so the fact that I sometimes talk as though Hutcheson were a hedonist and other times do not should be excused.

[6] Hutcheson holds that our ideas of primary qualities are only acquired by sensation in the sense that they are "concomitant ideas" with the purely sensible ideas, i.e. the secondary qualities. As Jensen notes, "Strictly speaking, Hutcheson's account of the primary qualities cannot be reconciled with his view that all of our ideas are gained from sensation and reflection. The primary qualities can be said to be gained from sensation only in the sense that they accompany sensations" (Jensen 1971, 41).

[7] In the *Inquiry*, Hutcheson says that "Grandeur and Novelty are two Ideas different from Beauty, which often recommend Objects to us" (*Inquiry* I.VI.xiii). This suggests that the internal sense gives rise to ideas besides the perception of beauty.

> Some strange Love of Simplicity in the Structure of human Nature, or Attachment to some favourite Hypothesis, has engag'd many Writers to pass over a great many Simple Perceptions, which we may find in our selves. We have got the Number Five fixed for our *external Senses,* tho Seven or Ten might as easily be defended. We have Multitudes of Perceptions which have no relation to any external Sensation; if by it we mean Perceptions, occasion'd by Motions or Impressions made on our Bodies, such as the Ideas of Number, Duration, Proportion, Virtue, Vice, Pleasures of Honour, of Congratulation; the Pains of Remorse, Shame, Sympathy, and many others. (EPA I.i, R 355)

The suggestion here seems to be that, because there many perceptions that are not given rise to through the powers of the external sense, there are many distinct kinds of perception of some internal sense. Because the perceptions of virtue and vice are among the perceptions listed here, it is natural to think of the perception of virtue and vice as the result of the operations of the internal sense. However, that this is not Hutcheson's meaning can be made clear from the rest of the passage:

> It were to be wish'd, that those who are at such Pains to prove a beloved Maxim, that "all Ideas arise from Sensation and Reflection," had so explain'd themselves, that none should take their meaning to be, that all our Ideas are either external Sensations, or reflex Acts upon external Sensations: Or if by Reflection they mean an inward Power of Perception, as I fancy they do, they had as carefully examin'd into the several kinds of internal Perceptions, as they have done into the external Sensations: that we might have seen whether the former be not as natural and necessary as the latter. (EPA Preface 5-6, R 355)

Hutcheson is here saying that the meaningful contrast is not between external and internal senses, but is, rather, between the external senses and the internal perceptions, or the "inward Powers of Perception."[8] On this reading, the internal sense is merely one of many distinct inward powers, as is the moral sense, and so the latter should not be thought of a species of the former.[9] The inner sense gives rise to a desire for the pleasures of the imagination.

---

[8] I am not suggesting that the difference is a difference between *sensation* and *reflection* as such. The thought is just that Hutcheson's term (albeit unfortunately misleading) is merely used to pick out one of several modes of internal perception.

[9] This interpretation differs from Jensen's who says that Hutcheson uses the terms "consciousness," "reflection," and "reflex sense," as interchangeable with "internal sense." See Jensen 1971, 41. I agree with Jensen's claim that the terms "consciousness," "reflection," and "reflex sense" are used interchangeably by Hutcheson to signify "the capacities by which the modifications of the mind, i.e. perceptions, judgments, reasonings, affections, and feelings are known as objects" (ibid.). My

3. Public sense: Hutcheson says that this sense is "our Determination to be pleased with the Happiness of others, and to be uneasy at their Misery" (EPA I.i). Hutcheson is clear that this sense is distinct from the moral sense insofar as people are usually very strongly affected by the "Fortunes of others, who seldom reflect upon Virtue, or Vice in themselves, or others, as an Object: as we may find in Natural Affection, Compassion, Friendship, or even general Benevolence to Mankind, which connect our Happiness or Pleasure with that of others, even when we are not reflecting upon our own temper, nor delighted with the Perception of our own Virtue" (ibid.).

The public sense gives rise to a desire for the happiness of others. Since Hutcheson takes all of the five primary classes of sense to give rise to original, or intrinsic, desires, we should think of the desire for the happiness of others that arises from the operation of the public sense as a desire for the happiness of others for its own sake. This is not to say that we do not sometimes desire the happiness of others instrumentally. What we see from the operation of the public sense is Hutcheson securing a fundamental place in his moral psychology for the intrinsic desires for the happiness of others. This point, as we will see below, is crucial for his arguments against the egoistic psychologies of Hobbes and Mandeville.

4. Moral sense: I will have more to say about this sense below. It is important to note that the moral sense is not what gives rise to an original desire for the happiness of others. The public sense is what Hutcheson claims gives rise to our benevolent affections. The moral sense gives rise to an original desire for virtue and aversion to vice (EPA I.ii).
5. Sense of honor: This sense responds to the perceived approbation or disapprobation of others towards oneself. Hutcheson says that this sense "makes the Approbation, or Gratitude of others, for any good Actions we have done, the necessary occasion of Pleasure; and their Dislike, Condemnation, or Resentment of Injuries done by us, the

---

disagreement with Jensen is just that I deny that the internal sense is identical with any of these terms. However, Hutcheson is not consistent in this usage throughout his corpus.

occasion of that uneasy Sensation called Shame, even when we fear no further evil from them" (EPA I.i, R 356).

It is important to keep in mind that Hutcheson's claim is that all *simple* ideas have their origin in one of the senses listed above. If, upon introspection, a simple idea is found that cannot be traced to (1)-(5), Hutcheson claims that another sense would then need to be posited to account for its origin. After describing (1)-(5), Hutcheson continues, "there are perhaps other perceptions distinct from all these classes" (EPA I.i, R 356). This positing of an additional sense should not be thought to be ontologically suspect: Hutcheson is merely describing certain psychological dispositions or "Natural Powers in the Human mind" (EPA I.i). By a simple idea, Hutcheson means an idea "which cannot be farther explained" (*Inquiry* intro., R 303).[10] Elsewhere Hutcheson makes it clear that it is not the case that cannot say *anything* more about such ideas; rather, whatever it is that we can say about them will not be particularly helpful. For instance, he says that such ideas "can only be explained by synonymous words, or by concomitant circumstances" (IMS 134, R 358). While Hutcheson never offers any arguments in support of the empiricist's project,[11] it is certainly the case that he takes the empiricist's project seriously and makes important contributions to its development, particularly with regard to those ideas that are important for moral philosophy: virtue, goodness, right, etc.[12]

---

[10] Hutcheson, Francis. 1725 *An Inquiry into the Original of Our Ideas of Beauty and Virtue*. 2nd ed. I will refer to both treatises of this text as *Inquiry* and citations will be to the Leidhold edited Liberty Fund editions (2008). When I need to refer to the first treatise, on beauty, I will include a Roman numeral 'I' before the section and article numbers. The Raphael (1991) anthology includes selections from the revised 4th edition, and I will include references to the marginal numbers in Raphael wherever possible.

[11] See Raphael 1947, 40 for an argument that the empiricist project is nowhere argued for, but is rather taken as an assumed starting point for Locke, Hutcheson, and Hume.

[12] Hutcheson often talks about there being "implanted in" human beings certain ultimate desires (most notably, an ultimate desire for the good of others). Is such a claim inconsistent with the rejection of innate ideas and principles so characteristic of the empiricist project? It is clear that Hutcheson does not think so, "We are not to imagine that this moral sense, more than the other senses, supposes any innate ideas, knowledge, or practical propositions" (*Inquiry* I.vii, R 313).

Again, Hutcheson holds that each sense gives rise to its own kind of original or intrinsic desire. He says, "Desires arise in our Mind, from the Frame of our Nature, upon Apprehension of Good or Evil in Objects, Actions, or Events, to obtain for our selves or others the agreeable Sensation, when the Object or Event is good; or to prevent the uneasy Sensation, when it is evil." (EPA I.i). It is through the external senses that we gain our ideas of things as being naturally good or evil. The various forms of inner perception all give rise in us to different kinds of pleasures and pains. Hutcheson thinks that we are so constituted that when we apprehend that something is good or pleasurable that we will come to desire it. Similarly, when we apprehend that something is evil or painful we will come to have an aversion to it. Original desires and aversions are the two fundamental *affections* and they arise only given the operation of some sense or other. The nature of the affections and their relationship to the senses is an important aspect of Hutcheson's theory to which I will return below.

Empiricism is the driving force for Hutcheson's arguments that our ideas of moral goodness and badness, of virtue or vice, and right and wrong all depend, in one way or another, on the operation of the moral sense.[13] The overall strategy for defending this theory is to argue that human beings have simple perceptions that cannot be accounted for without supposing a moral sense. Since these simple perceptions are present, and because they cannot be traced to the operation of any external sense or to some sort of internal reflection, Hutcheson argues that there must be some further sense.

Natural Goodness, Moral Goodness, and the Refutation of Egoism

In arguing for the existence of a unique moral sense, Hutcheson is not merely interested in refining the empiricist project as Locke advanced it. He is also concerned to argue against egoism and rationalism. Hutcheson's arguments for the existence of a unique moral sense are at

---

[13] As Radcliffe claims, Hutcheson defends the moral sense theory by appealing to "claims about the origin of ideas, including that of moral good and evil, in sense perception, and on observations about the phenomenology of moral judgment" (Radcliffe 2013, 10).

the same time arguments against the egoists.[14] The chief targets of this criticism are Hobbes and Mandeville.[15] Hutcheson, like many of his contemporaries, read Hobbes as endorsing a version of psychological egoism, but Hutcheson also seeks to defend the view that it is rational for us to act having as an end something besides our own interest. The argument, unsurprisingly, begins with an appeal to introspection. Hutcheson argues that there is a difference between those things that are called morally good and those things that are called naturally good or advantageous. Of the latter, Hutcheson says that the "pleasure in our sensible perceptions of any kind, gives us our first idea of *natural good* or *happiness*; and then all objects which are apt to excite this pleasure are called *immediately good*. Those objects which may procure others immediately pleasant, are called *advantageous*: and we pursue both kinds from a view of *interest*, or from *self-love*" (*Inquiry* intro., R. 304). This sense of good is distinguished from moral goodness, which "denotes our idea of some quality apprehended in actions, which procures approbation, attended with desire of the agent's happiness" (*Inquiry* intro., R 303).[16] Hutcheson's contention here is that people are affected differently by their perceptions of things as being naturally good on the one hand and their being morally good on the other:

> We are all then conscious of that *approbation* or perception of *moral excellence*, which *benevolence* excites toward the person in whom we observe it, and that opinion of *natural goodness*, which only raises *desire* of possession toward the good object. Now 'what should make this difference, if all approbation, or sense of good be from prospect of advantage?'… The reason why … must be this: 'that we have a distinct perception of *beauty* or *excellence* in the kind affections of rational agents;

---

[14] This is not to say that all of Hutcheson's arguments against the egoist are arguments for the existence of a moral sense. See Jensen's presentation of Hutcheson's *Essay* refutation of egoism in 1971, 14-22.

[15] As many commentators have noted, the full title of the first edition of the *Inquiry* announced that the treatise was one "In which the principles of the late Earl of Shaftesbury are explained and defended, against the author of the *Fable of the Bees*" (See Collected Works v.I "Bibliographical note"; Mautner 1993, 37; Jensen 1971, 14; 35). However, as Mautner and others have pointed out, Hutcheson removed this claim from the title page beginning with the second edition.

[16] Whether Hutcheson stays consistent with this definition throughout his corpus is unclear. In the *Essay*, Hutcheson presents a series of definitions and includes, "An action is morally good, when it flows from benevolent affection, or intention of absolute good to others" (Jensen 1971, 52).

> whence we are determined to admire and love such characters and persons.' (*Inquiry* I.i, R 307)

Hutcheson goes on to say that it is appropriate to think of this distinct perception of "beauty or excellence in the kind affections of rational agents" as the deliverance of a sense, that he calls a moral sense, "since the definition agrees to it" (*Inquiry* I.i, R 307). Hutcheson's arguments against the egoist turn on his appeal to introspection: "In this matter men must consult their own breasts" (*Inquiry* intro., R 303). His claim is that upon introspection, people will discover that they have different sentiments towards rational beings that act from an original desire for the happiness of others than they do towards that which is advantageous to themselves. He presents the following case: suppose that two people are equally advantageous to me. That is, both of them in fact succeed in getting me something in which I take pleasure. The first person acts out of a desire for my happiness for its own sake, while the second acts only from self-interest: "both are in this case equally beneficial or advantageous to us, and yet we shall have quite different sentiments of them" (*Inquiry* I.i, R 307).[17]

If all moral ideas and sentiments are reducible to self-love, or to considerations of natural goodness and advantage, then how can this distinct feeling of approbation towards those who act from an ultimate desire for our own good be explained: "What's Hecuba to us, or we to Hecuba?" (*Inquiry* I.ii, R 309). If self-love is what determines our approbation and condemnation in every case, then it would seem that our approbation would follow what we take to be advantageous to ourselves, and our condemnation would track what we take to be disadvantageous to ourselves. However, Hutcheson points out that people who are equally advantageous to ourselves or to others do not *ipso facto* equally determine our approbation. A related point in this debate is the fact that this distinct kind of approbation can be determined by the actions of people halfway across the world, by historical figures, and "toward even feigned

---

[17] This case, like many of Hutcheson's, is sadly under-described. It makes all the difference to the outcome of the case if the advantages I secure from these individuals are conceived to occur on one or two occasions, or if I am to rely on them in the future as well to secure my advantage. If the latter case is meant, then I have a very clear self-interested reason for not favoring the agent who acts from self-love, viz. his priorities could shift in those future cases where my advantage does not coincide with his own.

characters" (*Inquiry* I.ii, R 309). This approbation is not directly related to what is advantageous or naturally good for me, since the actions of these individuals may have nothing to do with my pleasure and pain. So, Hutcheson concludes, this approbation cannot stem from self-love.

This last line of reasoning may have been too quick. The egoist can respond that, while our approbation in these cases will not track the actual advantages we reap from such characters and their actions, we nevertheless can and do approve or condemn others regardless of place or time insofar as we apprehend that we have been benefited or harmed by similar characters.[18] Hutcheson's reply here is to point out that if this line of reasoning is correct, then it follows that our approbation should always be determined by what we apprehend would be most advantageous: "Would not the parsimony of a miser be as advantageous to his heir, as the generosity of a worthy man is to his friend? And cannot we as easily imagine ourselves heirs to misers, as the favourites of heroes? Why don't we then approve of both alike?" (*Inquiry* II.I.III, R 310). His answer: humans have a moral sense that "determines our approbation without regard to self-interest" (ibid.). Since we do not always approve of the side that in fact yields the greatest advantages,[19] and we do often condemn the victorious side of some conflicts, Hutcheson thinks that there must be some other internal principle responsible for determining our approbation, viz. the moral sense.[20] In this way, we can see that Hutcheson's rejection of egoism is merely the flip side of his arguments in favor of the existence of the moral sense.

---

[18] Hobbes, for instance, defines pity in *Elements of Law* as "imagination or fiction of future calamity to ourselves, proceeding from the sense of another man's calamity. But when it lighteth on such as we think have not deserved the same, the compassion is greater, because then there appeareth more probability that the same may happen to us. ..." (44). However, this is not Hobbes's only account of pity, and the account provided in *Leviathan* is less obviously self-regarding in this way. See Gert 1967, 509.

[19] It might be objected here that the approval of actions that *in fact* yield maximal self-advantage is an unfair interpretation of the egoist's position. However, even if we interpret the egoist as making a claim about actions that are likely to yield what is maximally self-advantageous, Hutcheson's response is still the same. It is not the case that our moral approbation always tracks those actions that we think are likely to yield maximal advantage to ourselves.

[20] Hutcheson considers two more responses that the egoist can appeal to in order to explain our pro-sentiments towards distant, historical or fictional figures. Since neither of these arguments furthers my aim here of connecting Hutcheson's rejection of egoism with his arguments for the moral sense theory, I will not consider them here. I mention these other arguments for two reasons. First, the foregoing may not

Some Brief Remarks on the Moral Ontology of the Moral Sense

Contemporary interest in Hutcheson's moral theory resurged for moral philosophers in the 1940's with the publication of D.D. Raphael's *The Moral Sense* and Norman Kemp Smith's *The Philosophy of David Hume*. Both noted the importance of Hutcheson's theory for the development of Hume's own. And, in the 1950's figures like Frankena, Peach, and Blackstone began to read Hutcheson as a figure that was developing a proto-noncognitivism. Hutcheson's theory, they hoped, might shed some light on how to overcome some of the more difficult aspects of that theory. The noncognitivist reading was soon contested, with commentators interpreting his metaethics variously as a version of moral realism, subjective descriptivism, and ideal spectator theory. In this section I will lay out some of the important claims that Hutcheson makes regarding the operation of the moral sense and will point to the various places where the interpretive controversy arises. While I do not take any of the arguments I offer in this section to be decisive, I offer some considerations against reading Hutcheson as a moral realist.

To begin, recall Hutcheson's definition of moral goodness from the *Inquiry.* He says that moral goodness "denotes our idea of some quality apprehended in actions, which procures approbation, attended with desire of the agent's happiness" (*Inquiry* intro., R 303). Jensen argues that this definition lends support to those who read Hutcheson as a moral realist because "[t]he very terms 'quality' and 'apprehended,' separated as they are from the term 'approbation,' lend themselves far more readily to a naively realistic interpretation."[21] The idea here seems to be that moral goodness is a property of actions that is detected by the moral sense, and subsequently gives rise to approbation and desire for the agent's happiness. The approbation and the desire for the agent's happiness are then psychological states that are the causal concomitants of the

---

be a satisfactory response to all readers and it's important to note that Hutcheson himself does not rest his entire case on this point. Second, in the next two responses, Hutcheson responds directly to views advocated by Hobbes and Mandeville, two of his biggest egoist opponents. So, there is some historical interest in these latter two arguments.

[21] Jensen 1971, 47-8.

apprehension of this quality of actions. On this interpretation, we approve of actions because they are good, and their goodness is entirely distinct from their being approved.[22]

However, it is not clear that the definition of moral goodness is as problematic for the anti-realist interpretation as Jensen seems to assume. The terms 'quality' and 'apprehended' are indeed distinguished from the term 'approbation' in such a way that a straightforward identification of moral goodness with approbation seems problematic. Nevertheless, the claim that "moral goodness….denotes our *idea* of some quality apprehended in actions" does not seem to suggest that Hutcheson identified approbation with moral goodness (emphasis added).[23] Moral goodness might still be identified with a distinct *psychological state* which gives rise to approbation and a desire for the happiness of the agent. Taken on its own, the definition of moral goodness is not as suggestive of the realist interpretation as Jensen makes it out to be.

However, this sort of tension in Hutcheson's formulation of his account of our perceptions of moral goodness is not infrequent throughout his corpus. For instance, he says, "[t]he quality approved by our moral sense is conceived to reside in the person approved, and to be a perfection and dignity in him" (*Inquiry* I.viii, R 314). This passage seems to suggest that the moral sense is a faculty that detects some objective quality of agents and it is this objective quality that Hutcheson identifies with moral goodness. It may be helpful to compare this to the

---

[22] A lot needs to be said in order to adequately address the debates between the realist and anti-realist readings. One of the major points of contention is how to distinguish realism from anti-realism. Since I do not intend to settle this dispute here, I will provide one of the accounts of moral realism that has emerged from this debate. I am not suggesting this is the most plausible or helpful account of the view. I rely on it because it will be helpful to see why the moral realist reading is problematic. Norton defends the view that "Hutcheson took the moral sense to be that principle of human nature which can and does enable us to apprehend and distinguish particular, objectively real features (intersubjective, independent of any particular observer) of the world about us, namely the moral features, with the consequence that we may be said to know the difference between vice and virtue, and to make moral judgments that are correct or incorrect, true or false" (1985, 416). As Winkler describes this view, it holds that "our perception of virtue and vice can only be compared to the perception of *primary* qualities, a kind of perception which is not merely a response to an objective world, but 'representative' or revelatory of the way things are" (1985, 180).

[23] Jensen presents this definition of moral goodness also on an earlier page where he did not include the crucial clause that moral goodness denotes an idea. He instead makes the unwarranted and misleading claim that Hutcheson says that moral goodness is a quality apprehended in actions. See Jensen 1971, 46.

operations of the external senses: just as the naïve realist about perception would maintain that our ideas of external world objects give us some reason to think that there are qualities external to our perceptions, so too does Hutcheson maintain that our perception of moral distinctions likewise maps onto qualities that are independent of our experience and these are mind-independent *moral* properties.

Hutcheson maintains that the "quality of actions" that gives rise to approbation and desire for the agent's happiness is benevolence. When we conceive of actions, either our own or other's, as being motivated by a desire for the well-being of others *for its own sake*, we experience a distinct sort of approbation and come to desire the happiness of the benevolently motivated agent. On the realist interpretation suggested above, this would seem to imply that moral goodness is to be identified with benevolence. The apprehension of an action as benevolent just is the apprehension of the action as being morally good because 'benevolently motivated' and 'morally good' are different ways of picking out one and the same property. On this view, if God had constituted human beings differently so that we did not approve of benevolent motives, it would still be the case that benevolence is morally good. We would just no longer approve of what is morally good.

It is not very likely that Hutcheson accepted this sort of a position. The reason is, briefly, that such a view draws too strong of an analogy between the moral sense and the external senses. Hutcheson is, however, quite clear that the moral sense operates not as an external sense, but as a form of inner perception. The moral sense is not a faculty of detecting empirical qualities of actions or agents, but is instead a faculty that gives rise to new (distinct) simple ideas upon contemplation of empirically and rationally given ideas.[24] I do not mean to suggest that

---

[24] This rejection of the naively realist reading is similar to the rejection offered by Jensen. However, Jensen's rejection places a heavy emphasis on some of Hutcheson's Lockean commitments so it's unlikely that a commentator like Norton would find them particularly compelling. Jensen says, for example, that "a naively realistic terminology in [Hutcheson's] moral sense doctrine conflicts with the fact that a naively realistic theory is entirely foreign to his general epistemology position" (Jensen 1971, 48). The general epistemological position that Jensen refers to here is one that denies the representationalism of the internal sense, on Hutcheson's view, largely because Locke repudiates such representationalism.

Hutcheson cannot be understood to maintain *some* version of moral realism, but instead that the version of moral realism that he could endorse must pay sufficient attention to the difference between the typical operations of the external senses and the modes of inner perception. The naïvely realist interpretation is not able to do this.

The anti-realist interpretations on offer all share a commitment to denying that Hutcheson thought of moral goodness as a mind-independent property. On such views, what is morally good is so only in virtue of certain subjective states of creatures constituted with a moral sense. Various readings that share this commitment go on to offer different accounts of the nature of moral judgments. Some commentators, spurred by Frankena's influential 1955 article "Hutcheson's Moral Sense Theory," have argued that Hutcheson's view about moral *judgments* is best understood as offering a rudimentary version of metaethical noncognitivism. While there are a few commentators who maintain that Hutcheson's view is best thought of as endorsing something like a version of noncognitivism, the arguments in favor of such a view are less than decisive.[25] The same should also be said of subjectivist interpretations of Hutcheson that maintain he held either a version of subjective descriptivism or a version of the ideal-spectator theory.[26] The sort of theory, be it realist, subjectivist or noncognitivist is left rather underdetermined by the textual evidence, though I suspect that the realist interpretation is the most difficult to reconcile with the whole of the text.

Hutcheson's Arguments Against the Rationalists

The remarks from the previous section are much too brief to settle anything with respect to Hutcheson's moral ontology. It is fairly clear that the moral sense operates by perceiving some quality in actions and responding with characteristic psychological states. I mentioned above that the quality of actions that Hutcheson finds essential to our moral ideas is benevolence.

---

[25] In addition to Frankena 1955, Blackstone 1965, Peach 1971, and Jensen 1971 all argue that Hutcheson's view is best thought of as a version of metaethical noncognitivism.

[26] Winkler 1985 maintains a subjectivist interpretation of Hutcheson.

Hutcheson takes settling this question to be one of the fundamental aims of his 1728 *Illustrations Upon the Moral Sense*: "1st, What quality in any action determines our election of it rather than the contrary? Or, if the mind determines itself, what motives or desires excite to an action rather than the contrary or rather than to the omission? 2nd, what quality determines our approbation of one action, rather than of the contrary action?" (IMS 133, R 358).[27]

What Hutcheson argues is that there are, broadly speaking, two ways of answering these questions: on the one hand, one may appeal to self-interest, like Hobbes, Mandeville, and other "Epicureans," and on the other hand, one may, like Hutcheson, maintain "that we have not only self-love, but benevolent affections also towards others…making us desire their happiness as an ultimate end, without any view to private happiness" (IMS 359). But Hutcheson argues that the rationalists mistakenly take there to be a third option. However, the overarching claim that Hutcheson makes throughout the *Illustrations* is that the rationalist's view, once properly understood, does not provide a theory distinct from either of these two:

> There have been many ways of speaking introduced, which seem to signify something different from both the former opinions. Such as these, that 'morality of actions consists in conformity to reason, or difformity from it:' that 'virtue is acting according to the absolute fitness and unfitness of things, or agreeably to the natures or relations of things,' and many others in different authors. (IMS Preface, R. 359)

One of the central claims that Hutcheson was trying to establish against the rationalists is that, far from giving an account of moral goodness, claims about the reasonableness of actions or an action's conformity to reason were hardly even intelligible. Such claims were prominent in the rationalist literature of the day. For example, William Wollaston, one of Hutcheson's frequent targets, presents the following definition of truth: "Those propositions are true, which express things as they are; or, truth is the conformity of those words or signs, by which things are

---

[27] Hutcheson, *An Essay on the Nature and Conduct of the Passion and Affections, with Illustrations on the Moral Sense*. 1728. Edited by Aaron Garrett. Indianapolis: Liberty Fund, 2002. When I need to refer to the *Essay* on its own I use the abbreviation 'EPA' and will use Roman numerals to indicate chapter and section number. When I need to refer to the *Illustrations* on its own, I use the abbreviation 'IMS' and follow with page numbers from the Liberty Fund edition. I will refer to both treatises together as *Essay and Illustrations*. I will provide references to the Raphael anthology where possible.

expressed, to the things themselves" (RND 274).[28] While Hutcheson would likely accept this characterization of truth,[29] he would reject the application that Wollaston makes of it: "I lay this down then as a fundamental maxim, that whoever acts as if things were so, or not so, doth by his acts declare, that they are so, or not so; as plainly as he could by words," and with respect to the moral worth of actions, "No act (whether word or deed) of any being…that interferes with any true proposition, or denies any thing to be as it is, can be right" (RND 279-80).[30]

Wollaston's view seems to be that no act is right which expresses a proposition that is inconsistent with the truth.[31] Wollaston provides a somewhat helpful example of the thought here. Imagine that a group of soldiers sees another group of soldiers approach.[32] The first group fires on the approaching group. Wollaston claims that the first group has, by this action, *declared,* "The approaching body of soldiers are our enemies." As it turns out, however, the approaching soldiers are not enemies at all, but auxiliary troops from their own side. Since firing on the second group was inconsistent with the truth, Wollaston would conclude that it was not right for the first group to fire.

---

[28] Wollaston, William. *The Religion of Nature Delineated.* Reprinted 1724 edition in Raphael's *British Moralists*. All Wollaston references will be to this text, henceforth RND, and will use the marginal pagination of Raphael 1991.

[29] As I discuss below, Hutcheson defines "reasonableness" as "Conformity to true propositions, or to Truth" (IMS 137, R 360).

[30] Wollaston is committed to the view that at least some actions are directly expressive of propositional content in the same way that declarative utterances are. On this view, there is a literal sense in which an act is either true or false. For more on Wollaston's account of declarative actions, see Feinberg 1977; Joynton 1981; Tweyman 1976; and Jensen 1971, 72-5.

[31] Joynton argues that, while Wollaston took falsehood to be both a necessary and sufficient condition for moral wrongness, he only offers an explicit argument for the sufficiency condition. See Joynton 1981, 440-1. It may also be helpful to note here that it is unclear whether Wollaston took *all* acts to be declarative (as Feinberg and Tweyman maintain) or whether he merely took *some* acts to be declarative of propositional content (as Joynton argues).

[32] Wollaston's case is under-specified. It makes all the difference how close to enemy territory they were, how much information they had about reinforcements, how hostile the overall war had been, etc. to evaluate the moral worth of the first group's actions.

As Hutcheson would undoubtedly understand this claim, it implies that the moral worth of an action depends on its being reasonable. This is so because of Hutcheson's definition of reason: "Since reason is understood to denote our power of finding out true propositions, reasonableness must denote the same thing, with conformity to true propositions, or to truth" (IMS 137, R 360). On Wollaston's view, rightness consists in conformity of actions to true propositions, and so rightness implies reasonableness.[33] What Hutcheson takes issue with here is that an action's being reasonable neither figures into our motives for action, nor is it sufficient to provide us with moral justification: "Reasonableness in action is a very common expression, but yet upon inquiry, it will appear very confused, whether we suppose it the motive to election, or the quality determining approbation" (ibid.). The complaint that Hutcheson here makes against the rationalist is that saying that some action is reasonable and meaning by that nothing more than that the action expresses, conforms to, or implies some true proposition is not sufficient for us to either have some motive to perform it, or to approve of the action. This is because there are true propositions about actions that are obviously vicious, so reasonableness, or conformity to truth, cannot be sufficient for rightness. In Wollaston's own example, the first group of soldiers was acting in conformity to the true claim "we are shooting our fellow soldiers." What makes the action-declaration of firing on one's fellow soldiers morally vicious, then, cannot merely consist in its being conformable to some true proposition.

What Hutcheson goes on to argue is that any criteria that could be added to the rationalist's conformity of actions to true propositions account must either allow for affections to play a fundamental role or the account must, after all, have recourse to a moral sense. So, for instance, Samuel Clarke maintains that there are "eternal and necessary differences of things" and that these "make it fit and reasonable for creatures so to act; they cause it to be their duty, or

---

[33] For my purposes here the differences between rightness/wrongness on the one hand and goodness/badness on the other, for Wollaston, will not amount to much since Wollaston says, "Moral good and evil are coincident with right and wrong. For that cannot be good, which is wrong; nor that evil, which is right" (RND 289).

lay an obligation upon them, so to do" (*Discourse* 225). Clarke then goes on to argue that understanding these eternal and necessary relations of fitness is, ordinarily, sufficient to determine one's will to act in accordance with them.[34] Of course, the will cannot be so determined in those agents whose understanding is depraved. Nor, Clarke adds, can those agents be moved whose "wills are corrupted by particular interests and affection, or swayed by some unreasonable and prevailing passion" (*Discourse* 230).[35] Our wills are determined by our understanding on this picture, and it seems as though the operation of affections or passions is nothing but a hindrance to acting in accordance with our duty. As we will see later, Hutcheson accepts that the *passions* do often interfere with acting morally. However, he rejects Clarke's view about the relationship between the *affections*, the understanding, and morality because he thinks that its attempts to ground morality in reasonableness alone are unsuccessful. Hutcheson argues that even the rationalist theory depends upon the operation of affections.

Hutcheson's arguments at this point appeal to a very important distinction that I have mentioned before, namely the distinction between, what he calls, exciting reasons and justifying reasons. I mentioned above that, as Darwall said, this distinction is roughly similar to, but not identical with, the distinction between explanatory or motivating reasons and justifying or good reasons for acting.[36] For Hutcheson, asking for a reason for an action may be asking "what truth shows a quality in the action, exciting the agent to do it?" (IMS 138, R 361). Here Hutcheson has in mind an explanatory or motivating reason. Hutcheson also notes that "sometimes for a reason of actions we show the truth expressing a quality, engaging our approbation" (ibid.). This sort of

---

[34] See Jensen 1971, 68-74 for a discussion of the ways in which a will can fail to be determined even when the agent understands the relations of fitness.

[35] It should be noted that Clarke was chiefly concerned with responding to Hobbes and other voluntarists. So, we should expect that his references to instincts or affections are all to self-interested instincts and affections. What he says about these, however, holds generally with respect to the source of our moral ideas. That is, even granting the existence of altruistic instincts and affections, Clarke will still deny that these altruistic instincts and affections are relevant to the account of the foundation of morals.

[36] Darwall 1992, 170 n8.

a reason is, for Hutcheson, a justifying or good reason.[37] However, Hutcheson's view seems to restrict what counts as a justifying reason to some true claim about a quality that engages our approbation. Thus, he is committed to a view about what counts as justifying moral claims. I will have more to say about his account of moral justification in what follows.

How does the distinction between exciting and justifying reasons help Hutcheson in his case against the rationalist? He claims that "all exciting reasons presuppose instincts and affections; and the justifying presuppose a moral sense" (ibid.). Insofar as the rationalist is committed to the view that reason alone can supply a motive to action, or that reason alone can justify the action independent of the operation of a moral sense, then, Hutcheson will argue that their view is unsuccessful.

Since Hutcheson's arguments for this point depend on his conception of the passions and the affections it will be helpful to say a little about said conceptions. As I discussed above, Hutcheson claims that there are several distinct sources of sensations. He distinguishes these from the passions and affections, and he characterizes the latter as "Perceptions of Pleasure or Pain, not directly raised by the Presence or Operation of the Event or Object, but by our Reflection upon, or Apprehension of their present or certainly future Existence; so that we are sure that the Object or Event will raise the direct Sensation in us" (EPA II.i).[38] What the passions and affections have in common is that they are perceptions of pleasure or pain that are not, unlike sensations, caused by the immediate perceptions of objects or events. The fundamental affections for Hutcheson are desire and aversion. Among the affections, Hutcheson includes joy, sorrow, and despair, as well as desire and aversion. Among the passions, which he

---

[37] This sense of 'justifying reason' is narrower than that provided in the introduction.

[38] It will be helpful throughout the following discussion to keep in mind that Hutcheson characterizes pleasure and pain as sensations. So, when he talks about "perceptions of Pleasure and Pain" he is not carelessly conflating sensation and perception, but is instead pointing to a kind of perception that takes the sensations of pleasure and pain as its objects.

sometimes calls appetites or instincts, Hutcheson includes hunger, thirst, anger, sexual drives, and the desire for companionship.[39]

There are two main differences between the affections, on the one hand, and the passions on the other. The first has to do with the different phenomenology that accompanies them. This difference is present in Hutcheson's formulation of the distinction by the 1st edition of the *Essay*, but becomes more pronounced by the 4th edition. The 1st edition account holds that a passion "includes" a confused sensation of pleasure or pain that is either caused by or accompanying "some violent bodily motions" (EPA II.i). Such passions are understood to interfere with "all deliberate Reasoning about our Conduct" (ibid.). In the 4th edition, Hutcheson says, "When the word passion is imagined to denote anything different from the affections, it includes a *strong brutal impulse of the will*, sometimes without any distinct notions of good, public or private, attended with a confused sensation either of pleasure or pain."[40] The violent bodily motions of the 1st edition have become "brutal impulses of will" by the 4th edition. So, as Jensen has noted, Hume's notion of the "calm passions" is straightforwardly contradictory on Hutcheson's account.[41]

The second main difference between the affections and the passions is that the former, unlike the latter, *only* arise when there is an apprehension, based on a previous perception, that the object or event in question will likely give rise to pleasure or pain. The idea here can be best illustrated by keeping in mind the sorts of things which Hutcheson lists as passions, *viz.* hunger, thirst, sexual drives, etc. We need not have any particular opinion that some object will cause us pleasure or pain in order to feel hungry or thirsty. Hutcheson does not deny that an apprehension of likely pleasure or pain may give rise in us to particular passions, but insofar as they function

---

[39] Jensen 1971,12.

[40] Quoted from ibid.; emphasis added.

[41] Ibid., 11. See chapter 5 for more discussion of Hume's theory of the calm passions.

in this way they do not act as *pure* passions. Instead, he refers to them as "passionate desires."[42] The passionate desires are contrasted with the calm desires. Hutcheson says, "There is a distinction to be observed on this Subject, between 'the calm Desire of Good, and Aversion to Evil, either selfish or publick [*sic*], as it appears to our Reason or Reflection; and the particular Passions towards Objects immediately presented to some Sense'" (EPA II.ii). Of these latter passions, the context makes clear that Hutcheson is not referring to pure passions since he takes himself to be addressing the issue of how "our Desires become passionate" (ibid.). The most important of the calm desires is calm universal benevolence. I will have more to say about this below.

To return to the role that the distinction between exciting and justifying reasons plays in Hutcheson's arguments against the rationalists, we have seen that Clarke seems to hold the view that understanding, via reason, the eternal and necessary relations of fitness determines one's will to act in accordance with them. Hutcheson thinks that this cannot be correct since every action presupposes the existence of an end and all ends presuppose some affection: "no end can be intended or desired previously to…self-love, self-hatred…benevolence toward others, or malice: all affections are included under these: no end can be previous to them all" (IMS 139, R 362). The exciting reasons we have presuppose that there are things that we desire, and these desires fall under these four categories of affections. The idea here is that an action is reasonable insofar as it is understood to be a means to some end the agent possesses in virtue of the operations of self-love, self-hatred, benevolence, or malice. Since we cannot go on desiring means to ends without presupposing some ultimate ends that are pursued for their own sake, Hutcheson argues that our ultimate ends are either desires for public or private good, or else are

---

[42] The 4th edition quote above characterizes passions as only occurring "*sometimes* without any distinct notions of good, public or private" which suggests that sometimes *passions* occur *with* distinct previous ideas of pleasure or pain from objects. I cannot find any place where Hutcheson uses 'passion' this way. So, I think it's best to read the "sometimes" as allowing Hutcheson to make room for the passions to operate together with the affections so that he can go on to distinguish passionate from calm desires and aversions.

aversions to public or private evil. It seems clear that one will only have a motivating reason to do some act when one believes that the act is a means to achieving what one desires.[43] Without the desired end, believing a true proposition about the act could not spur one into action. So, Hutcheson concludes that all motivating reasons presuppose affections.[44]

Hutcheson argues that there can be no exciting reason for our ultimate ends.[45] That is, for the desires of public or private good, or aversions to public or private evil, no exciting reason can be given. Hutcheson here appeals to Aristotle to make the point that there must be some things that are desired for their own sakes. Since exciting reasons are true claims about the means to our ends,[46] to say that we have an exciting reason for an ultimate end is just to deny that end was ultimate after all: "as to the ultimate ends, to suppose exciting reasons for them, would infer, that there is no ultimate end, but that we desire one thing for another in an infinite series" (IMS 139, R 362). Hutcheson gives a helpful illustration of this point. He has us imagine what a person who desires private happiness would say if we asked him what motivated him to pursue wealth. His response, Hutcheson suggests, is that wealth tends to bring about pleasure. If we continue by asking him what reason he has for pursuing pleasure, "one cannot imagine what proposition he could assign as his *exciting reason*. This proposition is indeed true, 'there is an instinct or desire fixed in his nature, determining him to pursue his happiness,' but it is not this reflection on his own nature, or this proposition which excites or determines him, but the instinct itself" (IMS

---

[43] In the *Essay*, Hutcheson lays out a set of axioms that he calls the "natural Laws of calm Desire" (EPA II.iv). Here he describes the ways in which desires arise from various beliefs about the degree and likelihood of concomitant and consequent pleasures and pains, measured by their intensity and duration, give rise to desires of varying strength.

[44] Hutcheson says that all motivating reasons presuppose either *instincts* or affections. He does not provide an explicit account of instincts, but it is clear from his use of the term that he takes instincts to be, or to be similar to, passions.

[45] As we'll see below, this might be too strong of a formulation of Hutcheson's doctrine since there is some sense in which reason can influence the selection of ultimate ends. See Darwall 1995, 225.

[46] For this formulation of Hutcheson's account of exciting reasons see Blackstone 1965, 13.

140).[47] Hutcheson provides a similar example to illustrate the same point with respect to disinterested reasons. He asks what reason a person has to risk her life in order to fight in a just war. He claims that this person's reasons will ultimately terminate in a disinterested affection for the happiness of mankind, or at least the happiness of other people in her country. For such an end, no exciting reasons can be given if her "affections be really disinterested" (ibid.).

Having argued that all exciting reasons presuppose either instinct or affection, Hutcheson then moves on to argue that all justifying reasons presuppose a moral sense. His argumentative strategy here is to present some of the standard claims that are thought to provide moral justification and show that they are either implausible or that they do require the operation of a sense and so cannot depend on reason alone. For instance, Hutcheson thinks that some of his rationalist opponents hold the view that an action is morally justified insofar as it is fit to attain some end.[48] To this position, he replies that many actions are fit to achieve some end, but not all of them have moral worth. He concludes that "justifying reasons…must be about the ends themselves, especially the ultimate ends" (IMS 145, R 363). Justifying reasons are about the ends of our actions. Given that Hutcheson conceives of justifying reasons as those true claims about qualities that determine our approbation, he asks of the rationalist, "Does a conformity to any truth make us approve an ultimate end, previously to any moral sense? For example, we approve pursuing the public good. For what reason? Or what is the truth for conformity to which we call it a reasonable end?" (ibid.). Hutcheson responds by claiming that there is no more reason to approve the pursuit of public good than there is reason for liking a pleasant fruit.

---

[47] I mentioned above that there might be some worry about the scope of Hutcheson's empiricism insofar as he often talks about implanted desires. This quote clearly shows that Hutcheson recognizes a distinction between desires and instincts, on the one hand, and propositions describing their objects or operations on the other. The rejection of innate principles is merely the rejection of the latter. There are no innate principles or propositions available to introspection describing the objects or operations of the desires that are in fact innate. For another interpretation along these lines see Paletta 2011, 151.

[48] While Hutcheson does use Clarke's language of "fit" here, this argument is not a successful objection to Clarke's view. This is because Clarke does not use "fit" and "unfit" in the sense that an action is fit or unfit to bring about some end.

As William Blackstone has noted, there is an ambiguity in Hutcheson's argument here. On the one hand, Hutcheson seems to be interested in arguing there are certain ends that are morally justified. That is, not only is it the case that we approve of promoting the public good, but our promotion of it, Hutcheson seems to argue, is morally justified. However, when Hutcheson formulates the question, "Does a conformity to any truth make us approve an ultimate end, previously to any moral sense?" he seems instead to be asking for a causal account. By asking what *makes* us approve an ultimate end, he seems to be signaling that he is interested in providing an account of the causes of our states of approval. As Blackstone suggests, there are, broadly, two ways to understand this ambiguity. First, one can hold that Hutcheson was inconsistent and that he at times speaks as if he were trying to provide justification while at other times he was speaking as if he were providing a causal account. Second, and perhaps more charitably, we might understand Hutcheson as arguing that there comes a point where theoretical reason is simply exhausted. On this interpretation, Hutcheson recognizes that all moral justification must eventually bottom out in the approval of the moral sense leaving this approval itself unjustified. There comes a point in our justifying our actions where no further reason can be supplied, and we only have recourse to the *causal* processes of a moral sense.[49] In a similar manner, our finding the taste of some fruit pleasant cannot be justified by appealing to some further feature of the fruit, such as it's being good. Rather, our justification reaches a point where all we can do is point to the operations of our sense of taste. So, Hutcheson concludes that the rationalist cannot maintain that understanding or reason, without the operation of a moral sense is ever sufficient to justify our actions. Our actions cannot be justified simply because we grasp certain truths about them: their justification depends on our having an ultimate approval of certain ends.

Against this latter claim, rationalists like Richard Price respond to Hutcheson by pointing out that his account of the operation of reason is artificially restricted. Hutcheson says,

---

[49] Blackstone 1965, 13-18.

"Approbation and condemnation are probably simple ideas" and, as we have seen, Hutcheson thinks that all simple ideas must be given by the operation of some sense and cannot be the deliverances of reason. Reason raises no new simple ideas. Price rejects this view, maintaining that "it is undeniable, that many of our ideas are derived from our INTUITION of the truth, or the discernment of the natures of things by the understanding. This, therefore, may be the source of our moral ideas" (*Review* 673). Since Price also holds that these moral ideas are simple ideas, what he is denying here is Hutcheson's contention that all simple ideas must come from the operations of a sense.[50] It must be granted that Hutcheson's arguments do nothing to impugn the more sophisticated rationalism that was developed in Price's intuitionism.[51] However, I do not take this to be a mark against his view. Hutcheson's aim in the *Illustrations* is to show that there are only two options when it comes to answering the dual questions he begins with: what qualities of actions motivate us to act and what qualities of actions engage our approbation? The answer to these questions either falls along broadly Hobbesian lines, or involves something like a moral sense theory. The rationalists of Hutcheson's day took themselves to be providing a different sort of answer: an action's conformity to reason provides an answer to both. However, their view was far from clear. So, while it certainly isn't the case that on *any* interpretation of the rationalist's view, it presupposes the operation of the moral sense, by pointing out that their view presupposed some basic operations of sentiments and affections, Hutcheson provided an

---

[50] Indeed, Price rejects the empiricist assumptions of Locke, Hutcheson, and Hume, arguing that their empiricist commitments are "destitute of all proof" (*Review* 674). All Price references will be to his *A Review of the Principal Questions in Morals*, henceforth '*Review*', and will use the marginal pagination of Rahpael 1991.

[51] Strasser argues that Hutcheson's refutation of rationalism is more successful than is usually thought. This is because Strasser seems to endorse the view that desires and affections for the happiness of oneself or others cannot be provided by reason alone. He notes that even Price must appeal to "natural biases" in addition to reason in order to account for these affections. See Strasser 1990, 55-61. Even if this is right, I do not think his interpretation bolsters Hutcheson as much as he assumes. This is because the operation of desire and affection is a different issue from the question of the origin of our understanding of the distinction between virtue and vice. Insofar as Hutcheson accepts the empiricist view that all simple ideas must have their origin in some sense, his arguments will not be able to respond to the sort of rationalist who, like Price, claims that the faculty of reason can give us simple ideas.

invaluable contribution to the debate. Jensen also notes this more optimistic reading of Hutcheson: "And, in fairness to Hutcheson, it might be remarked that some of the developments of the rationalists' theories in men such as Richard Price became more subtle largely through their having to meet Hutcheson's attacks on earlier rationalists such as Clarke and Wollaston."[52]

Passionate Desires, Calm Desires, and the Role of Reason

Hutcheson's claim that reason alone is not capable of moving us to action, and the role that this claim plays in his arguments that moral distinctions are perceived by sense and not reason is one of his most important and influential contributions to moral philosophy. Its influence on Hume is undeniable. Nevertheless, it would be negligent to misinterpret Hutcheson in such a way that reason plays no role in our moral lives. Indeed, Hutcheson is insistent that reason is crucial for the guidance of those affective and instinctive states he takes to be essential to our moral notions.

In order to see this, we have to return to the above discussion about the distinction between affections and passions. We noted above that, when the two terms are thought to refer to different ideas, that the latter, unlike the former, are attended with confused sensations of pleasure and pain and with violent bodily motions (I will refer to these two together as *bodily perturbations* in what follows). It is important to add to this observation Hutcheson's claim that these bodily perturbations interfere with the operation of our rational capacities. He says,

> When the word Passion is imagined to denote anything different from the Affections, it includes…a confused Sensation either of Pleasure or Pain, occasioned or attended by some violent bodily Motions, which keeps the Mind much employed upon the present Affair, to the exclusion of every thing else, and prolongs or strengthens the Affection sometimes to such a degree, as to prevent all deliberate Reasonings about our Conduct. (EPA II.i)

---

[52] Jensen 1971, 77. This interpretation is similar in some ways to an interpretation of Locke's arguments against innatism presented by Lowe. While many find Locke's arguments unsatisfactory, given the way the innatist was historically able to respond to them, Lowe argues that Locke's arguments are essential elements in the debate. The innatist's view to which Locke was responding was not very clearly formulated. By offering a set of objections, even objections that are unsuccessful insofar as they fail to exhaust the intellectual space available, Locke forced the innatist to specify their view. (Lowe, 31-2).

Insofar as the passions interfere with rational, practical deliberation, the notion of a rational motive for action must exclude the operations of the passions. This can be seen in Hutcheson's distinction, mentioned above, between passionate and calm desires.

Desires are affections and as such are always preceded by an apprehension of likely good arising from objects or events. When such desires arise independently from the phenomenological characteristics of the passions, that is, when they are free from bodily perturbations, they are calm desires. Calm desires can be directed towards any object that is apprehended as good, either for oneself or others. Since calm desires occur without the phenomenological characteristics of the passions, such desires present no obstacle to the deliberate use of our reason. In fact, Hutcheson characterizes the calm desires by their relation to reason when he says that there is a "Distinction to be observed on this Subject, between 'the calm Desire of Good, and Aversion to Evil, either selfish or publick [*sic*], as it appears to our Reason or Reflection; and the particular Passions towards Objects immediately presented to some sense'" (EPA II.ii, R 357). So, calm desires arise upon our apprehension of likely good to be obtained, either for ourselves or others, and since they occur through the use of reason, we should conclude that the reason-inhibiting characteristics of the passions are not present on these occasions.

Hutcheson recognizes that human psychology is more complex than this initial picture may suggest. Humans are susceptible to a wide variety of psychological "modifications" beyond those that are calm, rational and dispassionate on the one hand and those that are violent and opposed to reason on the other. Sometimes we have an apprehension of some likely good to be obtained and simultaneously undergo the phenomenological experiences characteristic of the passions. Here we have what Hutcheson calls a passionate desire.[53] It is rather strange that

---

[53] Hutcheson usually calls this category of mental modification passionate desire, but he does not think that this sort of mental phenomenon is restricted to *desires* alone. He can just as easily talk about passionate aversions. I have been presenting Hutcheson as holding that desire and aversion are both species of affections. However, he sometimes speaks as if 'desire' is synonymous with 'affection', as in the following: "[Desire and Aversion] are the proper Affections, distinct from all Sensation: *We may call*

Hutcheson chooses to describe these kinds of desires as passionate, since, as we have seen, one of the distinguishing features of the affections (desire and aversion) is that they, unlike the passions, arise from an apprehension of likely good or evil. He speaks of the passions as a "Propensity of Instinct to Objects and Actions, without any Conception of them as good." (EPA III.ii). Anger is a clear example of a *pure* passion. Hutcheson acknowledges that the typical occasions of anger may involve the apprehension of good to be achieved or evil to be avoided. One *desires* reparations for injury done to oneself as well as security from future injuries. But additionally, and independently from the apprehension of these goods, "there is in the passionate Person a Propensity to occasion Misery to the Offender, a Determination to Violence, even where there is no Intention of any good to be obtained, or Evil avoided by this Violence" (ibid.). Anger properly signifies only this propensity. "These propensities along with the Sensations above-mentioned, when they occur without rational Desire, we may call Passions, and when they happen along with Desires, denominate them passionate." (ibid.). As I suggested above, the way to interpret Hutcheson on this point is to notice that there are two characteristic differences between affections and passions. The passions do not require the apprehension of good or evil and are attended with violent, bodily perturbations. The affections arise only on the apprehension of good or evil and may or may not occur with violent, bodily perturbations. So, besides sensations, Hutcheson recognizes three broad categories of psychological modifications:

1. Pure passions: these arise independently from any consideration of likely good or evil and are attended with bodily perturbations[54];

---

both Desires if we please." (EPA III.ii; emphasis added). It is desire in this broad sense to which Hutcheson refers when he discusses passionate desires.

[54] Notice that Hutcheson will not allow for there to be something like a calm passion (Jensen notes that this is a category of Hume's which Hutcheson would find incomprehensible). The reason is that a calm passion would seem to involve a psychological modification that arises independently from any apprehension of likely good or evil *and* is not attended with bodily perturbations. Such a mental state would indeed be nonsense.

2. Calm affections: these arise only on the apprehension of likely good or evil and are not attended with bodily perturbations;
3. Passionate desires: these arise only on the apprehension of likely good or evil and are attended with bodily perturbations.[55]

Both Jensen and Darwall recognize that this threefold distinction allows Hutcheson to describe the mental modifications in terms of their varying degrees of susceptibility to reason. The pure passions are not directly susceptible to reason. The use of reason *alone* cannot strengthen or diminish the operations of the pure passions. However, the use of reason can indirectly affect the operation of the pure passions insofar as reason plays a role in the operation of the calm desires. It is the calm desires that are able to regulate or control the pure passions, and because the calm desires depend on reason, reason can indirectly regulate pure passions. Hutcheson says, "We obtain Command over the particular Passions, principally by strengthening the general Desires thro frequent Reflection, and making them habitual, so as to obtain Strength superior to the particular Passions." (EPA II.ii).

[55] It is worth pointing out, following Jensen, that Hutcheson takes the distinction between the pure passions and the calm desires to be a distinction in *kind* not in degree. The passionate desires are different in kind from the pure passions, given that they involve the apprehension of likely good. We should then think of the passionate desires are a particular species of desire with the calm desires picking out the limiting case of the degree to which such desires are attended with bodily perturbation. By contrast, Hume's roughly analogous distinction between the calm and the violent passions is a distinction in degree.

CHAPTER 4

INTERNALISM AND OBLIGATION IN HUTCHESON'S RESPONSE TO THE QUESTION "WHY BE MORAL?"

Introduction

In the previous chapter I laid out some of the fundamental elements of Hutcheson's moral sense theory as it relates to his moral psychology. I have yet to say anything about internalism or the sorts of answers that Hutcheson can provide to the question "why be moral?" It is important to have some ideas of the basic structure of Hutcheson's theory since it seems to provide some indication regarding the sort of internalist doctrine that Hutcheson could have accepted, while nevertheless very clearly ruling out any strict internalist interpretation. In this chapter, I take up the issues surrounding internalism in Hutcheson and his response to the amoralist's challenge.

Given that Hutcheson is constrained to offer an account that depends on contingent but non-arbitrary facts of human beings, it becomes rather difficult to maintain that he also accepted a version of internalism. We cannot simply read Hume's internalism, if indeed there is such a thing, back into Hutcheson's theory. Even though, according to Darwall, both Hutcheson and Hume fall within the *internalist* tradition that he calls empirical naturalism, I will argue, following a number of Hutcheson's interpreters (oddly enough, Darwall included), that Hutcheson cannot be thought of as an internalist in any strict sense.[1] This is because of his sharp distinction between affections and sensations, and his appeal to the exciting/justifying reasons distinction. I conclude this chapter by looking at the resources that Hutcheson's theory has to address the question "why be moral?" I argue that these reasons are grounded in interests that agents already have. If we reframe the question to ask about our reasons for expanding our circle

[1] Darwall argues that Hutcheson could have adopted a strict internalist theory of obligation, namely a version of the calm reflective deliberation version of empirical naturalist internalism. However, he claims that Hutcheson develops a sentimentalist theory that is not strictly speaking internalist. See Darwall 1995, 15-16; chapter 8. I will have more to say about this interpretive move in the concluding chapter.

of concern for others so that we will end up desiring the greatest good for the greatest number, Hutcheson can only point to non-moral, exciting reasons.

The Practicality of Ethics

It is often thought that moral philosophy is essentially practical. Moral philosophy is not merely concerned with understanding moral concepts like rightness and virtue and obligation, but is additionally concerned with how such concepts are to be translated into human actions.[2] Some moral philosophers adopt a particularly strong understanding of the practicality of moral philosophy. Some maintain that if a particular system of moral philosophy fails to have traction in the lives of those who defend it, we might well wonder whether there is something wrong with the system itself. In other words, if accepting the tenants a of particular scheme of morality does not inculcate in one a tendency to act in accordance with one's duty or to cultivate a certain kind of character, then the tenants of the scheme themselves are to be impugned. This conception of moral philosophy was well known and almost universally accepted when Hutcheson was writing.[3] Hutcheson is quite clear about his commitment to this fairly strong conception of practicality.[4] In his 1724 "Reflections on the Common Systems of Morality,"[5] Hutcheson states:

---

[2] Russell in "The Elements of Ethics" rejects this view. While he expresses dissatisfaction with some of the views expressed in this article, it is not clear that the rejection of the practicality of ethics is among them. See Russell 1970, 3. Russell's view here, like much in this article, is indebted to Moore who distinguishes ethics from the "much less respectable…study of casuistry" ([1902] 1988, 4). However, Moore claims that casuistry is "part of the ideal of ethical science," and adds that it "cannot be safely attempted at the beginning of our studies, but only at the end" (ibid., 5).

[3] See Mautner 1993, 32.

[4] For contrast, a fairly weak conception of the practicality of ethics might limit its claims about practicality to accepting some version of the 'ought implies can' principle. Hutcheson, it seems, would uphold the 'ought implies can' principle, but would maintain much else regarding practicality. For Hutcheson, it is not merely the case that the principles of moral philosophy should (theoretically) be such that we (normal, human agents) can practice them. Rather, for beings of our constitution, our study of these moral principles should (psychologically) incline us to act in accordance with them.

[5] The "Reflections on the Common Systems of Morality" [henceforth *Reflections*] was published in the *London Journal* to announce the upcoming publication of the first edition of Hutcheson's *An Inquiry Concerning the Original of our Ideas of Beauty and Virtue.* It can reasonably, then, be thought to give us a fairly good idea of the main concerns to which Hutcheson was to address himself in that book.

> 3. A very small acquaintance in the world may probably let us see, that we are not always to expect the greatest honour or virtue from those who have been most conversant in our modern schemes of morals. Nay, on the contrary, we may often find many, who have, with great attention and penetration, employed themselves in these studies, as capable of a cruel, or an ungrateful action, as any other persons: We shall often see them as backward to any thing that is generous, kind, compassionate; as careless of the interest of their country; as sparing of any expence; and as averse to undergo any danger for its defence, as those who have never made the law of nature their study: Nay, we shall often find them plentifully stored with nice distinctions to avoid their duty when it grows troublesome, and with subtile defences of some base practices, in which many an undisciplined mind would scorn to have been concerned. 4. Nor shall we observe any singular advantages arising from their studies, in the conduct of themselves, or in the state of their minds. We may often find them sour and morose in their deportment….Are all the efforts of humane wisdom, in an age which we think wonderfully improved, so entirely ineffectual in that affair, which is of the greatest importance to the happiness of mankind? Shall we lay it all upon a natural corruption in us, growing stronger, the more opposition it meets with? Or may we not rather suspect, that there must be some mistakes in the leading principles of the science; some wrong steps taken in our instruction, which make it so ineffectual for the end it professes to pursue?" (*Reflections* 96-7)

What we can see in this passage is that Hutcheson thinks that one of the major failings of the "leading principles" of moral philosophy is that they do not have an improving effect on those that study or defend them. There is nothing advantageous or ameliorative about accepting these schemes of moral philosophy: those that do are neither more likely to do their duty nor more likely to have agreeable personalities.[6] This line of criticism would have proved effective given that none of Hutcheson's contemporaries would have explicitly denied the practicality of moral philosophy in this sense.

It must be admitted that Hutcheson's discussion here is largely rhetorical. One wonders whether the tone of exaggeration has any sufficient reflection in reality: were *all* of Hutcheson's

---

[6] Mautner makes the same point in his Introduction to *Reflections* p. 31. Mautner is clear that the improving effect of moral philosophy ought to be seen in two areas. First, moral philosophy ought to make one morally better. Second, moral philosophy ought to make one "psychologically more harmonious" or "happier and more harmonious" (1993, 31). So, according to Mautner, the practicality of moral philosophy consists of three main aims: (1) moral philosophy should go beyond theory and be applied in practice; (2) moral philosophy should inculcate "psychological harmony or peace of mind"; and (3) moral philosophy should improve how people relate to each other (1993, 32). As I understand Mautner's position, the "should" involved in (1)-(3) is a should of theoretical rationality. (1)-(3) establish a set of standards for a successful account of moral principles.

peers such moral miscreants? Were corruption and cantankerousness such widespread concomitants of accepting egoism or rationalism? While answering such questions might yield a fascinating look at the secretly sordid lives of the famous British Moralists, such a line of inquiry misses the main point of the claim. Even if Hutcheson's peers weren't as depraved as he made them out to be, it is nevertheless clear that *if* they were so wretched, Hutcheson would place the blame, not on some fault of their individual psychologies or on a defect in human nature, but rather on a theoretical mistake in their moral philosophy.[7] Moral philosophy must have an improving effect on our relationships with others and on our own attitudes and character. Or, perhaps a weaker version of the thesis is more plausible: egoism and rationalism do not provide people with either enough motive or reason to do as morality requires or to be better, happier people. And to the extent these theories do not provide this they are conceptually flawed.

Another important problem with this passage as an objection to egoism and rationalism is that the egoist and the rationalist have a fairly straightforward response. If Hutcheson thinks their moral theories allow them to shirk their moral duties, then the obvious reply is that Hutcheson is mistaken about what those duties are. The egoist and the rationalist might just have different views about what is and what is not one's duty to perform. This isn't the sort of practical moral failing that Hutcheson has in mind here so it isn't clear that his objection is particularly compelling. I will have more to say about what I think Hutcheson can say to this response in my

---

[7] Hutcheson notes that this failure could be due to either some of the "leading principles" or to the manner of their "instruction." It seems he is leaving open that an individual could have been taught the correct principles of moral philosophy without them having an improving effect because their manner of instruction was problematic. This could very well have been a criticism of Joseph Butler, whose *Fifteen Sermons Preached at the Rolls Chapel* (1726) was to become a major influence on Hutcheson's own work. Hutcheson disapproved of preaching on subtle matters of moral philosophy in the pulpit. The role of the preacher is "not to explain the principles of the human mind, but to address himself to them and set them in motion" (Leecheman, quoted in Mautner 1993, 151). However, it is clear that the phrase "some wrong steps taken in our instruction" is ambiguous. On the one hand, it could mean that, while the moral principles according to which we were educated are correct, there was some mistake in the manner of our being taught them such that we were not practically influenced by them. On the other hand, it could simply mean wrong steps were taken in our education insofar as we were educated according to false moral principles. While the former interpretation may be correct, as evidenced by Hutcheson's later discussion of the proper role of the pulpit, I think that, given the immediately surrounding context, the latter interpretation is more likely.

concluding chapter. I argue there that part of what Hutcheson is thinking is that both the egoist and the rationalist are unduly focused on developing moral laws, precepts, and rules. Their theories always resolve themselves into a set of rule-based prescriptions. I think part of what Hutcheson is suggesting here is that this focus invites resistance. If moral prescriptions depend on self-interest, then any purported moral law must stand up to the sort of test proposed by Hobbes's fool. If moral rules depend on conformity to reason (whatever that amounts to), then it makes sense to ask why my will should conform to these rational dictates. So, even on the egoist's and the rationalist's own views about what one's moral duty is, Hutcheson is thinking that they will likely be "plentifully stored with nice distinctions to avoid their duty when it grows troublesome, and with subtile defences [*sic*] of some base practices, in which many an undisciplined mind would scorn to have been concerned" (ibid.).

If Hutcheson does indeed maintain this fairly strong thesis about the practicality of ethics, then his own theory should avoid the practical pitfalls of egoism and rationalism. That is, we should expect to find Hutcheson accounting for the superiority of his moral sense theory in its ability to foster practical ethical behavior. While many might say that the practical advantages of the moral sense theory over egoism are fairly clear, the superiority of Hutcheson's theory over rationalism is not so straightforward. Even though Hutcheson's theory creates a conceptual connection between morality and certain psychological states of agents, this by itself is insufficient to secure moral motivation. The rejection of rationalism, indeed, seems to hinge on this point: reason alone is incapable of generating the sorts of psychological states necessary for moral motivation and moral action. However, nothing in the moral sense theory guarantees that the psychological states necessary for moral action will be at all related to those psychological states that underlie moral perception and judgment.

In fact, Hutcheson's moral psychology in particular seems to preclude certain kinds of relationships between moral judgments and motivational states. As we have seen, Hutcheson is clear that the deliverances of the moral sense are sensations, and sensations by themselves have no motivational capacities. That I am determined to approve of those actions that I apprehend to

be motivated by benevolence does not *ipso facto* determine my motivational states with respect to that action. The mere fact that Hutcheson's moral epistemology gives an important role to sentiments does not by itself make it the case that his view is better able to accommodate moral motivation than either the egoist or the rationalist.

Obligation and Internalism

From what has been said so far concerning Hutcheson's moral psychology, it does seem strange to attribute to him a version of ethical internalism. As we have seen, Hutcheson argues against Price's view that understanding that a particular action is morally right will alone provide motivation to perform that action.[8] Since approbation and election are distinct, and the reasons determining one to the former need not determine one to the latter, we can have approbation with respect to actions that we have no motive to perform. Moreover, the operations of the moral sense give rise to sensations, and these are distinct from instincts, passions, and affections, or any motivational state recognized by Hutcheson. So, the operation of the moral sense does not directly raise any motivational states. It makes sense, then, why commentators have denied that Hutcheson holds a version of ethical internalism.[9]

As I noted in the introductory chapter, the term internalism is variously understood and applied. So, it is important to understand what it is that commentators are denying when they deny that Hutcheson held a version of internalism. Strasser seems to think that internalism is restricted to the view that the recognition of a moral truth is itself a source of motivation. This sounds like a version of non-constitutive motive internalism, which is commonly attributed to Plato, and to Price and other intuitionists.[10] Radcliffe, however, says that internalism is a view

---

[8] Strasser explicitly discusses this claim as a rejection of internalism. Strasser 1990, 64 n28.

[9] Strasser 1990, 64; Radcliffe 2013, 25 suggests that Hutcheson and Hume have different views on the rationality of moral action, in part, because of the latter's acceptance of internalism.

[10] This is the view that, necessarily, if an agent, S, actually epistemically engages with the fact that she ought to Φ, S will, in virtue of that epistemic engagement, thereby have some motivation to Φ.

that draws "a necessary connection between motivation and justification."[11] This formulation of the internalist thesis is compatible with all versions of motive internalism. Since Hutcheson argues that grasping a true proposition is insufficient for motivation, it seems clear that he will not accept the non-constitutive motive internalist view. Also, because the connection between motivation and justification is, strictly speaking, contingent, it is hard to see how Hutcheson could accept a version of the motive internalist thesis understood, as Radcliffe suggests, as a thesis involving a necessary connection between judgments and motives.

Jensen on Internalism

According to Jensen there is a more nuanced internalist thread in Hutcheson's account, and so, even though Jensen agrees with Strasser and Radcliffe that Hutcheson is *strictly* an externalist, he claims that "there is an air of internalism clinging to Hutcheson's theory."[12] He notes that among the many debts which Hutcheson owes Shaftesbury is that the former adopts from the latter a decidedly "internalistic mood" whereby moral actions are not conceived of as the result of egoistic and pragmatic calculations, but are rather seen "as shot through with intrinsic appeal."[13] Jensen argues that the air of internalist commitments surrounding Hutcheson's views can be seen in his rejection of egoism and rationalism. I will not recount these arguments here. What is important is that Hutcheson's rejection of egoism is different from his rejection of rationalism insofar as the latter theory, unlike the former theory, creates too wide a gulf between obligation and motivation. Hutcheson thinks that the error of the egoist lies *not* in their assumption that to have an obligation is to have a motive that is *in some sense reasonable*. Rather, the error of the egoist is that he fails to take into account other motivational states as

---

[11] Radcliffe p.10. In Radcliffe, 1996, she provides a more specific account of internalism when she argues that Hume accepts a particular version of appraiser internalism. See chapter 6 below.

[12] Jensen 1971, 103.

[13] Ibid., 67. The idea here, while not entirely clear, seems to be that, since morally right actions are "shot through with intrinsic appeal," when someone sees that she is obligated to do something, she will have some motive to do it *in virtue of its intrinsic appeal.*

elements of obligation. The egoist only recognizes self-interested obligation, but fails to take account of genuinely moral obligation.[14]

The distinction between interested obligation and moral obligation is found in both the *Inquiry* and the *Illustrations*, and can also be found in Clarke's *Discourse*.[15] In the characterization of interested obligation, Hutcheson agrees with two points of his egoist opponents. First, he agrees that there is a sense in which it is true of some agent that she ought to perform some action when she judges that by performing it she will promote her self-interests. Second, he agrees that to talk of obligation in this way is just to talk about a rational motive. He says,

> if, by obligation, we understand *a motive* from self-interest sufficient to determine all those who duly consider it, and pursue their own advantage wisely, to a certain course of action; we may have a sense of such an obligation, by reflecting on this determination of our nature to approve virtue, to be pleased and happy when we reflect upon our having done virtuous actions, and to be uneasy when we are conscious of having acted otherwise; and also by considering how much superior we esteem the happiness of virtue to any other enjoyment. (*Inquiry* VII.i, R 347; emphasis added)

The latter half of this quote will be important when we return to address the why be moral question below. For now, the important point is just that there is a sense of obligation, one of which we can become aware in ourselves by reflecting on the pleasures of acting virtuously, and

---

[14] It is worth pointing out that Hutcheson is critical of Shaftesbury's own view for the same reasons. As I discuss below, Shaftesbury held the view that, while one may approve of the actions of others for disinterested reasons, all rational motives must track the agent's own pleasures and pains. So, while Hutcheson does inherent the "internalist mood" from Shaftesbury, he nevertheless contends that Shaftesbury's account ignores the irreducibly moral sense of obligation.

[15] Darwall notes that the account of interested obligation is taken from Shaftesbury. But Hutcheson departs from Shaftesbury in maintaining that there is, besides this interested notion of obligation, a uniquely moral sense of obligation as well. This is important to note since it shows that Hutcheson is not merely concerned with responding to the egoists by distinguishing between two senses of obligation. He understands Shaftesbury to hold the view that, even though agents can approve of the actions of others without regard to self-interest, the only rational motive an agent has for virtuous action is that such actions promote the pleasure of the agent him or herself. So, by distinguishing the two senses of obligation, Darwall contends, Hutcheson is denying that the only rational motive to virtue is a motive of self-interest. Darwall 1985, 219-20.

that this sense of obligation is a motive that is reasonable in the sense that it specifies an exciting reason.

The relationship between moral, or disinterested, obligation and motivation is, however, less clear. This latter sort of obligation Hutcheson defines as "a determination, without regard to our own interest, to approve actions and to perform them; which determination shall also make us displeased with ourselves, and uneasy upon having acted contrary to it" (*Inquiry* VII.i, R 346). He adds, "this internal sense, and instinct of benevolence, will either influence our actions, or make us very uneasy and dissatisfied; and we shall be conscious, that we are in a base unhappy state, even without considering any law whatsoever, or any external advantages lost, or disadvantages impending from its sanctions" (ibid.). [16]

It is clear how the disinterested sense of obligation is meant to respond to what Hutcheson saw as the shortcomings of his egoist opponents. What is less clear is how exactly this sort of obligation is related to an agent's motivational states. Unlike the case of interested obligation, which Hutcheson clearly identifies as a motive, moral obligation is defined in terms of a "determination." Hutcheson elsewhere uses determination to refer to various kinds of perceptions, most notably in his definition of 'sense'. But it is clear that he wants the determination he refers to in his analysis of moral obligation to have some motivational role. He says it is a "determination…to perform" the actions approved of. He later specifies that "this internal sense, and instinct of benevolence, will…influence our actions" or leave us in a rather

---

[16] It is interesting to note that this formulation of moral obligation is from the 1738 revised 4th edition of the *Inquiry*. In the 1724 1st edition, Hutcheson says instead, "this moral sense, and instinct *toward* benevolence." Leidhold notes this change in the Liberty Fund edition, and Darwall uses it as an opportunity to note the development of the post-*Illustrations* account of obligation (Darwall 1995, 222). Darwall is right in pointing out that this change does make it clear that Hutcheson denies a direct motivational role for the moral sense. But it should be noted that in the 4th edition Hutcheson does not modify one of the more problematic passages in the *Inquiry*, namely his claim that moral obligation is a "determination…to approve actions, and *to perform them"* (VII.I, R 346, emphasis added). Darwall, as we will see, argues that the *Illustrations* account of obligation is entirely separated from notions of moral motivation. And if this is correct, one wonders why Hutcheson would not modify later editions of the *Inquiry* to reflect this.

miserable condition.[17] But now there is a bit of a problem. As we saw above, the moral sense does not play any direct motivational role. What's worse is that Hutcheson here seems to be suggesting that moral obligations depend, at least in part, on motives that an agent has in virtue of her having an aversion to being in a "base unhappy state." Such an account seems a bit too self-interested for Hutcheson's tastes.

This latter problem can be addressed by looking at the *Illustrations* definition of moral obligation: "That every Spectator, or he himself upon Reflection, must approve his Action, and disapprove his omitting it, if he considers fully all its Circumstances" (IMS 146). The approval or disapproval are not taken to be considerations additional to the moral obligation. That is, Hutcheson does not think that the pleasure of approval or the pain of condemnation are the only possible exciting reasons for one to keep one's moral obligations. He seems instead to be using them as the defining psychological states of this kind of obligation. If an agent, S, is morally obligated to Φ, then it is true of all spectators with a moral sense that know fully the circumstances surrounding S's Φ-ing that they will approve of S doing Φ and disapprove of S omitting Φ. The pleasures of self-approbation and the pains of self-condemnation are included in this account only insofar as they are causal consequences of S's having a moral sense and fully considering the circumstances surrounding her Φ-ing. While self-approbation and self-disapprobation may be part of what excites an agent to act, it does not follow that their absence removes the obligation. There is, then, no reason to think that Hutcheson's account of moral obligation is egoistic.[18]

---

[17] I argued above that Hutcheson seems to reserve the term 'internal sense' to refer to the sense of beauty in the *Essay*, but also noted that his use of it in this way is not consistent throughout his corpus. It is clear from the context that Hutcheson is using 'internal sense' to refer to the moral sense. This claim is supported by the comment that Hutcheson makes after presenting the corresponding definition of moral obligation in the *Illustrations*: he says of moral obligation that it "includes the moral sense" (IMS 146).

[18] Hutcheson sets as one of his fundamental aims in the *Inquiry* to establish that "what excites us to these Actions which we call Virtuous, is not an Intention to obtain even this sensible Pleasure…but an entirely different Principle of Action from Interest or Self-Love" (intro., R 306.). It would then be a rather unfortunate consequence if the only principle of action that Hutcheson succeeded in identifying were one that was not distinct from self-love.

With respect to the former problem, that the moral sense does not directly provide motivation, Hutcheson can appeal to the contingent relations between the senses and desires to explain how the moral sense, in approving of some actions, also determines us to perform them. Hutcheson does distinguish approval from election by noting that we may approve of actions that we ourselves could never perform. For instance, we can have approbation towards actions that were performed before we came into existence. As Strasser notes, this does not imply that our reasons for approval and the reasons that move to election can never coincide. When the reasons for approval are said to be reasons for election, there is implicit in these cases the claim that the agent could perform the action.[19] This allows for the possibility that the moral sense, the source of our reasons for approval, could in some cases serve as the source of our reasons for election.

It is important to note that such cases can only provide motivation indirectly and contingently. The motivation is indirect insofar as the moral sense only gives rise to sensations. Since these sensations are pleasures and pains, and Hutcheson maintains that desires follow the perception of pleasure and pain, the moral sense can indirectly give rise to desires. The motivation is contingent because Hutcheson denies that the indirect relationship between senses and desires is necessary. It just so happens that human beings are constituted in this way, but it could have been otherwise. Hutcheson says, "If it be here enquir'd, 'Could not the Deity have given us a different or contrary determination of Mind, viz. to approve Actions upon another Foundation than Benevolence?' It is certain, there is nothing in this surpassing the natural Power of the Deity" (*Inquiry* VII.xii).[20]

Because of this, Jensen notes that, in a strict sense, Hutcheson's view is a version of externalism: "the relationship between moral judgments and connation must be causal, as in the

---

[19] Strasser 1990, 90.

[20] Hutcheson argues that, because God did create us in such a way that the moral sense approves of those actions that tend towards the greatest happiness, this is some evidence for God's goodness. So, while God could have constituted us in such a way to make us approve of "Barbarity, Cruelty, and Fraud" (ibid.), that he did not do so gives us some probabilistic reason for concluding that God is good.

case of 'externalist' theories, not logical, as in the case of 'internalist' theories."[21] However, Jensen goes on to argue that Hutcheson's account is best viewed as a sort of middle ground between two extreme accounts of the relationship between moral judgment and motivation. On the one hand, there are those accounts that seek to identify moral judgment with motivational states, so that it becomes a straightforward contradiction to assert that S is obligated to Φ but has no motive to do so. On the other hand, there are those accounts that seek to provide only a very weak causal connection between moral judgment and motivation. Since Hutcheson maintains that the relationship between approval and desire is contingent, he does not maintain the former position. However, it would be wrong to think of him as holding the latter position since he holds that the causal connection between approval and desire is quite strong. As Jensen says, the connection between the two "is as close as it may be in the realm of matters of fact."[22]

Strictly speaking, then, the indirect motivational capacity of the moral sense is a contingent feature of human nature. But Hutcheson does argue that our psychological constitution is not entirely arbitrary. He asks, "Why may not the Deity have something of a superior kind, analogous to our moral sense, essential to him?" (IMS 153, R 366).[23] Insofar as God has created us with a moral sense, and has given us desires that arise on the perception of pleasures, it isn't the case that humans just arbitrarily approve of and desire whatsoever they may choose. Of course, Hutcheson maintains that there could be creatures that approve and desire differently from ourselves. But when we are considering purely human affairs, it is *as if* there is a necessary connection between our states of approval and our states of desire.

---

[21] Jensen 1971, 98. It should be noted, however, that the mere fact that there is no logical connection between moral judgments and moral motives is not sufficient to deny that any version of internalism obtains. For instance, non-constitutive internalists maintain that moral facts, moral judgments and moral motives are all distinct, and are all merely causally related when S makes a true judgment that she ought to Φ. What makes this a version of internalism is that the causal connection between S's judgment that she ought to Φ and S's motivational states is a necessary one.

[22] Jensen 1971, 98-9.

[23] Jensen notes that this argument was not altogether satisfying to Hutcheson, and that it opened him up to criticism when we consider the further question whether what God approves of is morally good. Jensen p.100.

Darwall on Internalism

Darwall too notes the difficulty in maintaining that Hutcheson held a version of internalism. He argues for an interpretation of the *Inquiry* account of obligation that is similar to the one provided by Jensen. However, Darwall argues that this account undergoes significant revision in the *Illustrations*. The most important revision is that the account of moral obligation does not make any reference to an agent's motive. The *Illustration* account of moral obligation, again, is that "every Spectator, of he himself upon Reflection, must approve his Action, and disapprove his omitting it, if he considers fully all its Circumstances" (IMS 146). The distinctive sense of moral obligation depends entirely on the sentimental responses of observers. Given the strong separation between sensation and motivation outlined in the *Essay*, it seems as if Hutcheson is distancing himself from the *Inquiry*'s claim that moral obligation involves a "determination…to perform" the actions of which the moral sense approves. So, Darwall says, the *Illustrations* account of moral obligation has "not even an implicit reference to motive."[24]

Darwall finds this shift in Hutcheson rather surprising. He claims that Hutcheson's account of obligation in the *Illustrations* is a *sentimentalist* account, and since Hutcheson distinguishes so sharply between sentiments and motives, and because he accepts the *calm deliberation view* of practical reason, we might have expected him to argue that moral obligation consists in an agent's reason-informed motives. He says that Hutcheson's account of obligation is "an 'internal' theory, but not an internalist theory in our sense, because it situates the moral 'ought' within the perspective of an *observer* rather than that of a deliberating *agent*."[25] Since Hutcheson could have taken the sort of internalist view open to Cumberland, Darwall finds this sentimentalist approach to *moral obligation* to be "somewhat puzzling."[26]

---

[24] Darwall 1995, 233.

[25] Ibid., 234.

[26] Ibid. I will have more to say about Darwall's finding this view puzzling in chapter 7.

What is the alternative view of obligation that Darwall thinks is open to Hutcheson? In order to see this, we need to understand what Darwall calls the calm deliberation view of practical reason.[27] This view is meant to contrast with the kind of account of practical reason that is adopted by Hobbes. Recall that, for Hobbes, reason is a purely theoretical faculty. That is, reason is not thought to include any operations that are not involved in empirical inquiry (broadly construed to allow reasoning about necessary truths). On Hobbes's account, practical reason does not provide agents with ends but serves only to direct action in such a way that agents can pursue ends that they have already. As a theoretical faculty, reason is able to tell us that some actions are means to particular ends. But in order for reason to *dictate* actions to agents, it must reveal these actions as a way of achieving ends that the agent who is deliberating about what to do already has. Given that agents have as an inescapable end their own preservation, reason directs or dictates to agents the means through which their preservation can best be achieved, namely the laws of nature. This is the instrumentalist view of practical reason.

Hutcheson shares with Hobbes the view that reason is purely a theoretical faculty. As Darwall expresses this idea at one point, reason has "no end or practical 'ought' intrinsic to it."[28] It may also seem to be the case that Hutcheson accepts the instrumentalist account of practical reason. Recall, for instance, the account of exciting reasons that Hutcheson provides in the *Illustrations* as a "truth [that] shews [*sic*] a quality in the action, exciting the agent to do it…" (IMS 138, R 361). As Hutcheson understands this account, he seems to think that one can have an exciting reason to Φ only if Φ is in the service of other ends that one already possesses. He argues, for instance, that there can be no exciting reasons for our ultimate ends, "but as to the ultimate ends, to suppose exciting reasons for them, would infer, that there is no ultimate end,

---

[27] Elsewhere, Darwall calls this the calm deliberation view of "practical normativity" (See Darwall 1997, 84). I mentioned above that this view is found in Cumberland with whom Hutcheson, according to Darwall, shares a great number of commitments. Considerations of Cumberland's theory would take us too far afield and so will not be pursued in this dissertation.

[28] Darwall 1995, 209-10.

but that we desire one thing for another in an infinite series" (IMS 139, R 361). The idea here seems to be that reason discovers that certain actions have qualities that engage our election of them, but when it comes to our ultimate ends, the things we pursue for their own sakes, reason is unable to discover any qualities that determine our election of certain ends over others. That is, there are no exciting reasons for us to desire some things as ends: our ends are entirely given by the present constitution of our instincts and affections. What this suggests is that while reason plays no role in the determination of ends, reason is useful in determining those actions that bring about the ends we already have. In other words, Hutcheson seems to endorse the instrumental view of practical reason.[29]

However, Hutcheson does seem to accept a view of practical reason that places a stronger emphasis on the role of reason than the instrumental account will allow. Throughout the *Essay*, Hutcheson develops an account of moral psychology in which reason, in addition to its instrumentality, plays a very significant role. Hutcheson draws a distinction between the calm affections and the passionate desires, and this distinction makes it clear that reason does much more than simply direct us to our antecedently given ends. Reason reveals to us ends that, given our psychological make-up, we cannot but adopt. One way of seeing this is to look at Hutcheson's discussion of calm universal benevolence. This is a desire for the well-being of the most extensive system of sentient beings for its own sake. He recognizes that many people have desires for the good of some more limited system of sentient beings (like one's country or some political faction thereof), and for the good of particular others (friends, family, etc.). But he notes that there are people who nevertheless lack calm universal benevolence. This happens when people do not make reflective use of their reason. If they had, Hutcheson maintains that they

---

[29] As Darwall argues in "Hutcheson on Practical Reason," the argument here seems to conflate two separate issues. On the one hand, there is the overall claim that all exciting reasons presuppose instincts and affections. On the other hand, there is the claim that all exciting reasons presuppose some further end, such that there can be no exciting reasons for ultimate ends. As Darwall notes, it is possible for all exciting reasons to presuppose instincts and affections even where there are exciting reasons for pursuing an ultimate end. See Darwall 1997, 81.

would come to desire the well-being of the most extensive system of sentient beings for its own sake.

Hutcheson, setting the stage for Hume, maintains that reason by itself is entirely inert in human action. Reason is nothing other than our power of finding out true propositions. Now, some of these propositions concern qualities of actions that motivate us to perform them, namely the exciting reasons. So, in virtue of the fact that we come to have desires for things that we apprehend, though reason, to be good, reason has an indirect effect on our motivational states. But this doesn't mean that reason has no role in determining the ends that agents have. As Darwall has argued, there are some things that reason can reveal to human beings and, given our psychological constitution, we can come to adopt new ultimate ends. In order to make this argument, Darwall relies on two of Hutcheson's claims: (i) the apprehension of good tends to give rise in us to a desire for that which we think is pleasurable, either for ourselves or others, and (ii) the desire does not arise in virtue of its being a means to some further end that we already have, like a desire for the greatest pleasure. Hutcheson says, "we need not imagine any Innate idea of Good in general, of infinite Good, or of the greatest Aggregate: Much less need we suppose any actual Inclination toward any of these, as the Cause or Spring of all particular Desires" (EPA II.ii).

How does this relate to the adoption of new ends? Darwall says, "[Hutcheson] holds that we are naturally so constituted that when we apprehend the prospect of pleasure [natural goodness] we tend to desire it, to seek it as end."[30] Darwall seems to mean that when we judge that some action or object Φ will give rise to pleasure, for ourselves or others, we then come to desire Φ. Keep in mind that what Darwall is arguing for here is the view that the "discovery of some truths (through reason) can move an agent to adopt something as an ultimate end."[31] What this seems to suggest is that we do not desire pleasure (natural goodness) as an ultimate end, at

---

[30] Darwall 1995, 225.

[31] Ibid.

least not until we judge that something is pleasurable. What is true is that we judge certain things to be good, and then tend to desire them. But this does not mean that we have an abstract ultimate desire for good in virtue of which this particular desire arises as a means.

Since Hutcheson also maintains that people do not usually have *calm universal* benevolence, and since he maintains that the prospect of pleasure gives rise to desire, he seems to suggest that there is a way to acquire calm universal benevolence and this will involve the use of reason. This would mean that people come to have an end, the well-being of the most extensive system, that they did not possess prior to the operation of reason. He says, for instance, "Reflection may actually intend universal absolute Good; but with the common rate of Men their Virtue consists in intending and pursuing particular absolute Good, not inconsistent with universal Good" (EPA II.ii, Law 12).[32] Moreover he claims, "Our moral Sense, tho it approves all particular kind Affection or Passion, as well as calm particular Benevolence abstractedly considered; yet it also approves the Restraint or Limitation of all particular Affections or Passions, by the calm universal Benevolence. To make this Desire prevalent above all particular Affections, is the only sure way to obtain constant Self-Approbation"(EPA II.ii). So, when Darwall claims that it is puzzling that Hutcheson offers a sentimentalist account of obligation, what he seems to be suggesting is that Hutcheson could have easily accepted an account of obligation that appealed, not to the sentimental responses of observers, but to the calm, reason-informed motives of agents. According to Darwall, "Hutcheson implicitly rejected without any evident rationale the understanding of obligatory force that had become so widely accepted in early modern British ethics (especially among empiricist-minded philosophers)."[33]

---

[32] Hutcheson contrasts absolute good with relative good. Something is absolutely good when the amount of pleasure it contains or produces is greater than the amount of pain contained or produced. Something is relatively good when the pains are not outweighed by the pleasures. The claim that people intend particular absolute good that is "not inconsistent with universal good" is ambiguous between "not inconsistent with *actual* universal good," "not inconsistent with *probable* universal good," and "not inconsistent with *reasonably foreseeable* universal good." Given the other laws of calm desires, Hutcheson makes it clear that it is the latter interpretation that matters. See Darwall 1995, 227-8; 1997, 84-5.

[33] Darwall 1995, 234.

For Darwall, what Hutcheson's view in the *Illustrations* amounts to is that human agents are driven by "two independent and incommensurable sources of *rational* motivation."[34] Hutcheson maintains that our ultimate rational ends are a calm self-love and a calm universal benevolence. The sense in which both of these ends are rational is that they are our ultimate exciting reasons. And although reason can raise a calm universal benevolence in those agents that have only particular passions and affections, Hutcheson does not seem to argue that such benevolence should be a stronger influence than self-love when the two are thought to conflict. Darwall cites two important passages from the *Essay and Illustrations* to support this claim: "the Strength of either of the private or publick [*sic*] Desire of any Event, is proportioned to the imagined Quantity of Good, which will arise from it to the Agent, or the Person for whose sake it is desired" (EPA II.iv); and "When any Event may affect both the Agent and others, if the Agent have both Self-Love and publick Affections, he acts according to that Affection which is strongest, when there is any Opposition of Interests" (IMS 143). These passages suggest that there is no in principle reason why calm universal benevolence should be thought to always provide a stronger motive for an agent to act than self-love.

Darwall's argument here is that Hutcheson "must think the proportion to be different for the two cases, private and public."[35] Hutcheson says that the strength of a desire is proportioned to the imagined quantity of good. This suggests that the greater the amount of imagined good the stronger the desire. When Darwall says that Hutcheson thinks the proportion is different in the case of private and pubic desires, he seems to be saying that Hutcheson thinks the same amount of good will produce different strengths of desire depending on whether the desire is private or public. He claims that Hutcheson must have held this view because of the account of universal benevolence. This account maintains that universal benevolence "always includes the agent's

---

[34] Ibid.; emphasis added. Darwall credits Hutcheson with starting a line of thinking in British moral philosophy that extends to Sidgwick's dualism of practical reason.

[35] Ibid., 235.

good as one among others."[36] Darwall argues that if the amount of good produced were all that determines the strength of the desires, then Hutcheson would have said that benevolence is the over-riding motive in those cases where self-love and benevolence conflict. Since acts motivated by benevolence always include one's own good, such acts will always have a greater imagined amount of good and so would generate stronger desires. Darwall argues that Hutcheson does not say this but instead says that in cases of apparent conflict between self-love and benevolence, the agent will act on whichever desire is strongest. So, Hutcheson must think that the proportion of good to the strength of the desire is different in the case of private desires than it is for public desires.

Finally, Darwall notes that, for Hutcheson, there never really is any opposition between self-love and universal benevolence. This is because God has created us in such a way that we derive the greatest pleasures from reflecting on our benevolently motivated actions. He quotes: "if he discovers this Truth, that 'his constant pursuit of the publick Good is the most probable way of promoting his own Happiness,' then his Pursuit is truly reasonable and constant; thus both Affections…are at once gratified, and he is consistent with himself" (IMS 143). Self-consistency, here taken to be a matter of gratifying our two ultimate ends, is clearly something desirable according to Hutcheson. This self-consistency is made possible in us only in virtue of God having created us in such a way that we approve of, and get the most pleasure from, those actions which tend to the greatest happiness of others. The idea here seems to be that, though it may appear from our perspective as agent's deliberating about what to do that there is often a conflict between self-interest and universal benevolence, there really is no opposition. God ensures that the pursuit of public good is the most probable means to achieving one's own happiness.

Darwall's discussion of the two incommensurable sources of rational motivation is problematic. First, Darwall argues that Hutcheson was committed to the view that the proportion

---

[36] Ibid., 236.

of good to the strength of the desire is different in the case of private desires than it is for public desires. When agents perceive a conflict between these motives, then, there is no deliberative perspective from which they determine which one is more reasonable (understood in terms of a question about exciting reasons). But this view does not seem to fit with other claims that Hutcheson makes. For instance, Hutcheson says that the claim that the pursuit of public good is the most probable means for achieving one's own happiness is "an *exciting Reason* to *serve the publick Interest*, since this Conduct is the most *effective Means* to obtain both ends" (IMS 144, R 362). This does not suggest that self-love and benevolence are incommensurable motives at all. Recognizing that pursuit of public good is the most probable means to one's own happiness is "absolutely necessary" for agents that have both self-interested and benevolent motives (IMS 144). Without such recognition, an agent "does not act *reasonably* for his own Happiness, but follows it by *Means* not tending effectually to this *End*" (IMS 143). One's own happiness cannot be secured in ignorance of this claim because whenever one acts for the sake of public good he will think he is neglecting his own self-interest. Similarly, whenever he acts for his own self-interest, he will think that he is neglecting the public good (IMS 143-4). These claims suggest, contrary to Darwall's proposal, that Hutcheson does think that benevolence is always the stronger motive, at least when agents understand the connection between private and public goods.

Secondly, we might well wonder whether the incommensurability arises with respect to our justifying reasons for acting. I think it is important to make clear that it does not. Hutcheson says that if we are asking whether the end of public good is more reasonable than the end of private good, and we mean by this "does not every spectator approve the pursuit of public good more than private," he thinks "the answer is obvious, that he does" (IMS 144, R 362). It is tempting to look at Hutcheson's discussion of justifying and exciting reasons as independent theses about the nature of rationality and action. However, it is important to keep in mind the argumentative contexts in which they arise. This, as I will argue in the concluding chapter, is one of the chief difficulties arising for Darwall's reading of Hutcheson. The passages from the

Illustrations that are under discussion arise as a part of Hutcheson's multi-faceted argument against the rationalists. The overall claim that he is trying to establish is that there can be no exciting reasons independent of instincts and affections, and there can be no justifying reasons independent of a moral sense. So, when Hutcheson says that there is more justifying reason to prefer pursuit of public good over pursuit of private good, his main point is that this justifying reason is not established "for any *Reason* or *Truth*, but from a *moral Sense*" (IMS 144, R 362). Keeping the argumentative context in mind, these passages do seem to suggest that there is more reason, both exciting and justifying, to pursue public goods over private.

Why Be Moral?

I turn now to make explicit the resources that Hutcheson has to be able to address the question "why be moral?" I first present an interpretation offered by Douglas Paletta. I argue that part of his account is roughly correct insofar as he traces the exciting reasons to be moral to the satisfaction of interests that agent's already have. However, Paletta's way of developing this account is problematic. I then present and evaluate his claim that Hutcheson needs to, and does, respond to a different kind of amoralist challenge. In the second section, I present Radcliffe's interpretation of the "why be moral" question in Hutcheson. I argue that her discussion can be improved by noting how Hutcheson thinks two elements of her interpretation are related.

Paletta's Two Skeptical Challenges

Paletta argues that there are two important challenges that moral skepticism poses to Hutcheson's theory. First, there is what Paletta calls the "skeptic's challenge": what reason does an individual have to be moral? Second, there is what Paletta calls the "priority challenge": what accounts for the greater weight that is thought to attach to moral considerations over other considerations?[37] The above discussion of Darwall's interpretation suggests that Hutcheson

[37] Paletta 2011, 150.

already meets this latter challenge. However, since both Darwall and Paletta seem to think the issue is worth discussing, I think it will be helpful to continue to examine it in what follows.

Paletta frames his interpretation as a discussion about how Hutcheson's account of moral motivation can respond to this pair of challenges. He argues that Hutcheson responds to the skeptic's challenge by drawing an analogy between the reasons why individuals should pursue natural goods, and the reasons why individuals should pursue moral goods. Recall that Hutcheson defines natural goodness in terms of pleasure and natural evil in terms of pain, and then characterizes those objects that raise such sensations as being immediately naturally good. Hutcheson adds, "Those Objects which may procure others immediately pleasant, are call'd Advantageous: and we pursue both Kinds from a View of Interest, or from Self-Love" (*Inquiry* intro., R 304). As Paletta reads this, it implies that "natural goods are those things that give us pleasure. Determining what gives us pleasure in turn establishes what is in our interest."[38] So, the exciting reason that an individual has for pursuing the natural goods is that such goods promote her happiness, which is something about which she is concerned to the extent that she loves herself.[39] On this interpretation, an agent's motivational states with respect to natural goods track what's in her interest.

Paletta argues that Hutcheson could respond to the moral skeptical challenge by offering an analogical line of reasoning with respect to the moral goods. Why should an individual be moral? Acting morally, on Paletta's interpretation, means acting for the sake of those things that bring about the greatest happiness for the greatest number. What reason does one have for promoting the happiness of others? One has a reason for pursuing the happiness of others insofar as one has public affections, including benevolence, and so can be said to love others. Hutcheson

---

[38] Ibid., 151.

[39] Hutcheson seems to think that everyone does in fact love him or herself, and seems skeptical that there is anything like an ultimate desire for one's own unhappiness: when he points to this as one of the possible motivating affections he says, "desire of private misery (*if this be possible*)" (IMS139, R 362; emphasis added).

holds that people do in fact have this other-directed love or esteem. So, individuals have reason to act on benevolent motives.[40] On this interpretation, one has reason to act as morality requires insofar as doing so brings about the satisfaction of one's, in this case other-directed, desires.[41]

There are a couple of problems with this parallel account. First, Hutcheson defines moral goodness as an idea of a property of an action that determines our approbation independently from any consideration of self-interest. Paletta notes that Hutcheson must then go on to specify what this quality is that determines our approbation in this way. Paletta is, however, less than clear on what he thinks Hutcheson's account of this property is. He says that the property is something like the action's being useful to mankind. This suggests that our approbation is determined by our understanding the actual usefulness of actions to the general good. However, Paletta adds, "rather than self love, which seeks to advantage the agent, *benevolence* motivates moral acts by aiming to promote the general good."[42] This suggests that the quality that determines our approbation isn't the action's actual usefulness, but its being motivated by benevolence, which in turn aims at, but does not necessarily achieve, the promotion of general good.

What's more problematic for Paletta's argument is his appeal to the notion of moral goods that are analogous to natural goods. Natural goods can be the objects of motivational states for Hutcheson. Motivational states do not require natural goods, since humans can be moved by passions and these do not require an apprehension of likely (natural) good or evil. Moral goods, whatever these are supposed to be, cannot be the objects of motivational states in the same way, since all motivational states presuppose natural goods as ultimate ends.[43] Suppose that an agent

---

[40] Paletta 2011, 154.

[41] Paletta does not clearly distinguish exciting from justifying reasons for action. I think it is fairly clear that he means to be talking about exciting reasons when he develops his parallel account of reasons to pursue natural goods and reasons to pursue moral goods. However, as I argue presently, his view runs into problems because it isn't always clear that he is only talking about exciting reasons.

[42] Ibid., 151; emphasis added.

[43] Darwall 1995, 221-2, 229.

apprehends that some action open to her is morally good. This implies that the agent experiences approbation upon contemplation of her performing this action. This in turn implies that the action she is contemplating is one that would be motivated, at least in part, by benevolence which has as its end some natural good. That is, if she thinks the action is morally good, then she must approve of herself doing it from a desire for the well-being, understood in terms of natural goods, of others for its own sake. Her perception that the action is morally good is motivationally irrelevant insofar as whatever additional motives she might acquire through the use of the moral faculty (e.g., the pleasures of reflective self-approbation) would not add anything to the moral worth of the action. Moreover, these additional motives would only have her own natural good as their end.[44] Given the different relationships between natural goods, moral goods, and motivational states, I do not think a successful account of Hutcheson's response to the skeptic's challenge can depend on an analogy between them.

These issues aside, on Paletta's interpretation both accounts of the reasons that an individual has for pursuing natural goods and the account of the reasons that an individual has for pursing the moral goods appeal to the desires and drives that an individual already has: self-love in the former case, and other-directed affections or benevolence in the latter. This is an important feature of Hutcheson's account of the practicality of ethics. Hutcheson cannot appeal to an agent's grasping some general truth about morality or about human nature as the ground of a motivating reason for everyone. He says,

> This proposition is indeed true, "There is an Instinct or Desire fixed in his Nature, determining him to pursue his Happiness;" but it is not this Reflection on his own Nature, or the Proposition which excites or determines him, but the Instinct itself. This is a Truth, "Rhubarb strengthens the stomach:" But 'tis not a Proposition which strengthens the Stomach, but the Quality in the Medicine. The Effect is not produced by Propositions shewing [*sic*] the Cause, but by the Cause itself. (IMS 140)

---

[44] As we will see in chapters 6 and 7, Hume seems to accept a similar claim that a regard to the moral worth of an action cannot provide a motive to perform it in the same way that a regard to its ability to produce pleasure can.

As we saw above, Hutcheson denies that reason, independent from instinct and affection, can provide motivation. His account of moral motivation will then appeal to reasons that are relative to the psychological makeup of the individual in question. Radcliffe points out that this idea is developed in the course of Hutcheson's correspondence with rationalist Gilbert Burnet. Hutcheson begins:

> To a being which acts only for its own happiness, that end is reasonable which contains a greater happiness than any other which it could pursue; and when such a being satisfies itself with a smaller good for itself while a greater is in its power, it pursues an unreasonable end … But if there are any beings which by the very frame of their nature desire the good of a community or which are determined by all kind affections to study the good of others and have withal a moral sense which causes them necessarily to approve such conduct in themselves or others…, to such beings that end is reasonable which contains the greatest aggregate of public happiness which an agent can procure;…If these beings also have self-love…and at the same time find that their highest happiness does necessarily arise from kind affections and benevolent actions, that end which would appear reasonable would be universal happiness…, for both desires are at once gratified as far as they are capable of doing it by their own actions. (Corr., 209-10)[45]

This shows, as Radcliffe points out, that the psychological states of the agent determine what counts as a reasonable end for that agent.[46] Both Paletta and Radcliffe, then, agree that Hutcheson has a response to the skeptical challenge that appeals to the present interests of the agent in question. As Paletta says, "by defending the claim that we do act benevolently, Hutcheson defends the empirical claim needed to explain why someone should act morally."[47] On both interpretations, Hutcheson would say that one should be moral, i.e. act from benevolent motives, because one in fact has other-directed desires and these desires can best be fulfilled by acting from benevolent motives.

---

[45] *The Correspondence between Gilbert Burnet and Francis Hutcheson* in *Illustrations on the Moral Sense* ed. Bernard Peach 1971. References to this text will be abbreviated "Corr.' and will be followed by the page number of the Peach edition.

[46] Radcliffe 2013, 13. Radcliffe calls this a relativistic conception of reasons. The idea is that there is no such thing as a rational course of action for agents to pursue *simpliciter.* To say that a course of action is rational is always to say that that course of action is rational *for so-and-so.*

[47] Paletta 2011, 154.

We ought here to be reminded of Darwall's interpretation, since this account cannot adjudicate between the views of self-love or benevolence. As Radcliffe notes, "no actions are reasonable in themselves, and [Hutcheson's] theory allows the possibility that there *could* be beings who simply have no reasons to behave morally, but we are not among them."[48] Hutcheson thinks that all people "probably" have a moral sense.[49] On Radcliffe's interpretation, it seems, the moral sense is what gives an agent their other-directed affections. If an individual lacked the moral sense, Hutcheson would have nothing to say to provide such a person with a reason why he ought to behave as morality requires.[50] Paletta, too, notes a similar consideration. But, for him, this raises a problem for Hutcheson, namely the priority challenge. He says that by offering an account that draws a strong analogy between the exciting reasons agents have to pursue pleasure and the exciting reasons they have to be moral, Hutcheson places moral reasons "on the same footing with the reasons to act for self interest."[51] As a consequence of this, there seems to be no further exciting reason to pursue what is moral than there is to pursue what is advantageous to myself. So, by appealing to the desires that agents happen to have, Hutcheson does not find an in principle way to say that moral reasons have a special weight or priority over

---

[48] Radcliffe 2013, 15.

[49] *Illustrations* p.132 (Peach edition). Later in the same paragraph, Hutcheson says, "We conclude that all men have the same affections and sense." This suggests that he does think all people have a moral sense and have benevolent affections. He may have qualified the former statement to signal that the moral sense, just like every other sense, may be missing or defective in some persons. Just as one might be blind, one's moral sense may also be dysfunctional. But the defects of the moral sense are harder to detect than the defects of the external senses, and so by saying that all men "probably" have a moral sense, Hutcheson may just be signaling that we may find people that are "morally blind."

[50] This account, however, is not quite correct. As we saw above, the moral sense gives rise to a desire for virtue, but does not give rise to benevolent motives themselves. This point is obscured in the correspondence with Burnet. At some points, Hutcheson seems to talk about the purely self-interested agent as a being that merely lacks a moral sense. At other points, he suggests that the purely self-interested agent lacks *both* a moral sense *and* other directed desires. Radcliffe does not discuss this ambiguity, but rather leaves the case undecided. She says that Hutcheson "argues that apart from having *a moral sense or benevolent affection*, he can see no explanation why anyone would approve of actions that pursue public good over private interest" (Radcliffe 14, emphasis added). I will have more to say about this issue in the final section of this chapter.

[51] Paletta 2011, 154.

the reasons deriving from self-interest. Without such methods of adjudication, Paletta argues, Hutcheson is unable to fully address the dual challenges of moral skepticism.

So, how does Paletta attempt to resolve the issue? First, he notes that Hutcheson's theory of moral motivation does not entail that self-interested and benevolent motives can never be simultaneously present in an agent. On the contrary, Hutcheson explicitly allows that agents may be motivated to act by both self-love and benevolence on the same occasion. He says,

> If a man have such strong benevolence, as would have produced an action without any views of self interest; that such a man has also in view private advantage, along with public Good, as the effect of his action, does no way diminish the benevolence of the action. When he would not have produced so much public good, had it not been for prospect of self interest, then the effect of self love is to be deducted, and his benevolence is proportioned to the remainder of good, which pure benevolence would have produced. (*Inquiry* II.iii, R 319)

What is important for Paletta's argument is the first case that Hutcheson mentions, where an action is causally over-determined by both self-love and benevolence, and where there would have been just as much natural good (for others) produced if self-love hadn't figured into the account at all. According to Paletta, Hutcheson took benevolence to be of more weight than self-love in these cases of mixed motivations, because observers would call the action benevolent even though self-love is present. Paletta has moved from talking about the strength of exciting reasons to talking about the relative weight that observers place on different motives to determine if an action is, all things considered, self-interested or benevolent. He says that in "pure cases…interest approves of self loving acts similarly to how the moral sense approves of benevolent acts," but he argues, "in mixed cases benevolence has more weight than self love."[52] The idea seems to be that if mixed cases were like pure cases, then it should be a draw how to evaluate the action. However, Hutcheson claims that we evaluate the action as benevolently motivated, which suggests that the *moral sense* offered the weightier evaluative reason in this case. So, Paletta argues, Hutcheson can appeal to the weightier moral concerns to solve the priority challenge: our evaluative reasons favor moral motives to those of self-interest.

---

[52] Ibid., 157.

This line of argumentation is not successful. First, Paletta seems to misunderstand Hutcheson's claims about causally over-determined actions. In the passage from the *Inquiry* quoted immediately above, Hutcheson nowhere mentions that benevolent motives are given more evaluative weight than other motives. His concern is primarily with those actions that are not casually over-determined in the sense that is relevant for Paletta's argument. Hutcheson is mainly concerned with those actions where the agent would not have brought about as much public good were it not for his additional motives of self-interest. In these cases, it's not that the benevolence has more weight than the motives of self-love. Rather, the moral worth of the action overall is to be understood only in terms of the good that would have been produced independent of self-love. This is because the moral sense approves of benevolently motivated actions, not self-interested ones. By contrast, in the cases of mixed motives where the same action would have come about by either motive independently, but it just so happens that the action was jointly determined, it is still not clear that Paletta is entitled to his claim that there is special weight to the moral motives. All that Hutcheson says is that the benevolence of such an action is not diminished by the mere fact that the action was jointly motivated by self-love.

Secondly, as I mentioned above, Paletta moves from talking about the exciting reasons for action to talking about the relative weight that observers place on these motives when they are assessing the all things considered motive of the action. Hutcheson clearly holds that motives come in varying degrees of strength and that they are related to quantities of foreseeable goodness, but this doesn't seem to be what Paletta is after when he talks about relative deliberative weights. What is most problematic about this shift is that it forces Paletta to begin talking about what the moral sense most approves of. His account, which has all along been about Hutcheson's theory of moral *motivation*, turns into a discussion about Hutcheson's theory of moral *justification*. Once we see that the "moral evaluation" involved in assessing the all things considered motive for a particular action arises from the moral sense, Paletta's conclusion becomes rather unhelpful: "All things equal, the asymmetry in the approval of mixed actions provides some reason to give considerations of benevolence more weight than considerations of

self love because the moral sense approves of these actions more strongly than interest."[53] If the reading of the *Inquiry* passage that I suggested in the previous paragraph is correct, then Hutcheson is not saying benevolence is given a special weight. Paletta's conclusion, then, becomes the claim that the moral sense approves of benevolent motives. That spectators approve of benevolent motives in no way suggests that benevolent motives are to be given some special deliberative weight when assessing the all things considered motives of an action. Further, it is trivial on Hutcheson's view that benevolent motives are given a "special weight" when morally evaluating an action: benevolent motives are the only motives relevant for assessing moral worth.

Finally, I am not convinced that Hutcheson needs to solve the priority challenge. As I argued in response to Darwall's reading, Hutcheson seems to think that he can explain why pursuing the public good over one's own private good is more reasonable in both the exciting and justificatory senses. Paletta says, "according to the priority challenge, an account should try to do more than establish that we have some reason to be moral. The account should further explain why pursuing moral goods tends to have priority over pursuing natural goods."[54] If we understand the pursuit of moral goods as having priority over the pursuit of natural goods as the former pursuit being more reasonable than the latter, then Hutcheson's discussion at IMS 144-3 (R 362) seems to provide just such an account.

Radcliffe's Proposals and Their Connection

Radcliffe points out that Hutcheson discusses three different ways in which he could respond to the question "why be moral" where this question is understood to be a question about *motivating* reasons.[55] While she points out that Hutcheson discusses the three possible responses, Radcliffe notes that he explicitly rejects the third. The three possible responses are:

---

[53] Ibid., 158.

[54] Ibid., 143.

[55] Radcliffe 2013, 12.

1. An agent has a reason to be moral because acting from benevolence fulfills interests that she already has;
2. An agent has a reason to be moral in virtue of the fact that reflective self-approbation triggered by her benevolently motivated acts will give rise in her to pleasure;
3. An agent is motivated to be moral because her justifying reasons serve equally as her motivating reasons.

The third possibility is rejected because of Hutcheson's distinction between exciting and justifying reasons, and so I will not spend time on it here. I will focus my discussion on the first and the second possible responses to the question "why be moral?" I have briefly presented Radcliffe's discussion of (1) above, but will here develop the account in more detail. Then I argue that (1) and (2) should be thought to work together in ways that Radcliffe could, but does not, identify.

As we have seen, the first response maintains that agents have reasons to be moral because acting benevolently fulfills interests that they already have. This account is obviously non-egoistic: if benevolently motivated actions are what will fulfill my interests, then I obviously have interests in the happiness of others for its own sake, otherwise I would not have benevolent motives. And, given that Hutcheson does maintain that humans are in fact benevolently motivated, we can satisfy our interests in the happiness of others for its own sake by acting on motivational states that we already possess.[56] This is similar to Paletta's claim that by claiming that we do act benevolently, Hutcheson establishes the empirical basis for the claim that we ought to act benevolently. Or, to put it in Radcliffe's words, we have reasons to be good insofar as doing so fulfills our interests.

---

[56] It may be worth noting that Radcliffe's evidence for the claim that human agents are in fact benevolently motivated is that she, correctly, reads Hutcheson as maintaining that all human agents have, or likely have, a moral sense. This might seem like insufficient support since the moral sense gives rise to approbation, and Hutcheson is clear that benevolence and approbation are distinct. Elsewhere, however, Radcliffe offers an interesting interpretation of Hutcheson's account of love, and this interpretation makes it plausible to suppose that the moral sense not only gives rise to approbation but, in some manner that Hutcheson does not make explicit, also give rise to benevolence. See Radcliffe 2004.

Recall that Hutcheson maintains that there are no exciting reasons for ultimate ends, and that ultimate ends are determined entirely by the affective states of the agents for whom they are ends. This means that a question about some end being more reasonable than another must always only refer to the affective constitution of the person evaluating the ends: there is nothing further to which an agent could appeal in order to settle the issue. If a being did not have a moral sense (and so did not approve of benevolence) and also did not have public affection (and so was not him or herself benevolently motivated), then Hutcheson does not see in what sense such a being could ever maintain that pursuit of public good over pursuit of private interests is reasonable for her. However, Hutcheson does maintain that human beings have both a moral sense and benevolent motives: in response to a potential objection from his rationalist opponents who say that it is absolutely reasonable for a being that does no harm to others not to be subjected to pain, or that benevolence is absolutely reasonable, Hutcheson says, "It is very probable that every man would say that these things are reasonable. But then all mankind have this moral sense and public affections" (Corr., 211). So, because of our constitution we will find it reasonable that, for example, "smaller private good should yield to greater public good" (Corr., 210). There are, then, for Hutcheson, no ends (or actions in the pursuit of ends) that are reasonable *simpliciter*, but all ends and actions are reasonable only insofar as they are *reasonable-for-so-and-so*.

I have not raised any difficulty for Radcliffe's interpretation, nor do I intend to. I do not wish to argue that Radcliffe is incorrect: quite the contrary, for the reasons stated above, I think Radcliffe's interpretation is preferable to that offered by Paletta. However, I do think that there is an important respect in which Radcliffe's account perhaps does not go far enough. We can see this by first asking what Hutcheson has in mind when he says that all human beings have public affections.

I think there is good reason to read 'public affections' as, if not identical to benevolence, then at least including benevolence as a species of public affections. In either case, what we find is that these affective states are rather multifarious depending on their objects and their strength.

What I wish to draw attention to here is that, even if Hutcheson does recognize that all human beings have these public affections, it does not follow that all people have the public affections that underlie morally praiseworthy action in all cases. Given that benevolence is not a singular phenomenon according to Hutcheson, the mere fact alone that all humans have benevolence *of one sort or another* does not mean that such motives will direct all humans towards moral action. In order to develop this thesis, there are two issues that I need to address. The first is defending why I think Hutcheson maintains that benevolence, or public affection in general, is multifarious. The second is to elaborate on the claim that a benevolently motivated agent might not be moved towards moral actions.

While there is much to say about the first issue (that benevolence is multifarious), I will keep my comments here brief. There is straightforward textual support for this claim throughout Hutcheson's corpus. The title of Section V of the *Inquiry* is called "A further Confirmation that we have practical Dispositions to Virtue implanted in our Nature; with a further Explication of our Instinct to Benevolence in its various degrees…," and in this section Hutcheson says right off the bat that "we are not to imagine that this Benevolence [which is found in all human beings] is equal, or in the same degree toward all. There are some nearer and stronger Degrees of Benevolence, when the Objects stand in some nearer relations to our selves" (*Inquiry* V.i, R 341). The stronger degrees of benevolent motives, then, are directed at more limited systems of rational beings like our family, friends, and neighbors.

So what are we to make of this fact that human beings have benevolent motives that naturally come in varying strengths? It should be fairly clear that this fact does not undermine the consideration that, again, benevolent agents will find it reasonable for a smaller private good to be set aside in order to pursue a greater public good. However, what is unclear is whether such agents will find it reasonable that a public good, perhaps of more limited scope, should yield to a wider public good. In the *Essay*, Hutcheson makes sure to distinguish our calm public desires as they are directed towards "the good of *particular Persons* or *Societies* presented to our Senses; or that of some more abstracted or general Community, such as a *Species* or System" (EPA II.ii,

R 357). This is the distinction between universal calm benevolence and calm particular benevolence and it is only a quantitative distinction. Moreover, there seems to be no guarantee that beings that have particular benevolence will by that fact alone have universal benevolence.[57] The worry here is not that agents, while having the capacity for benevolence, never in fact realize this motive in their action because it is always thwarted by stronger self-interested passions. The concern that I am raising here is presupposing that agents have benevolence, and that their acting on this motive is not thwarted by particular self-interested passions. That is, I am only here concerned with benevolence insofar as it is a calm desire for the well being of others. The question that I am raising is simply: what others?

To add a further complication, recall that Hutcheson remarks that if there is a conflict between the public desires and the private desires, or even if it appears to an agent that there is such a conflict, then "that kind prevails which is *stronger* or more intense" (EPA II.ii). This suggests that the amount of good is not the only relevant feature in determining an agent's will. This claim is part of what leads Darwall to suggest that self-interest and benevolence are incommensurable. And while Paletta does not explicitly say so, I think this claim leads him to think that Hutcheson needs to address the priority challenge.[58] I maintain that this claim raises a different concern for Hutcheson. This concern arises when there is, or is perceived to be, a conflict between public desires of different scopes. Where there is, or where there is perceived to be, a conflict between particular benevolence and universal benevolence, then that kind prevails which is the most intense. Since public affections are strongest when directed towards family members, friends, and neighbors, it seems that the motive of particular benevolence will be the one that prevails.[59]

---

[57] I henceforth leave off the 'calm' qualifier. I think that the influence of the passions on our reasoning makes it fairly clear that passionate desire and the pure passions will not provide us with the sort of reasons that are relevant here.

[58] Paletta cites Darwall's interpretation as posing the question that his own response to the priority challenge is meant to address. Paletta 2011, 157.

[59] I will discuss a similar concern for Hume in chapter 6.

I have shown that Hutcheson holds the view that benevolence is not a singular phenomenon, but is multifarious. I have also shown that the strength of one's benevolence depends on its objects. So, since Hutcheson does have a view according to which the reasonableness of an action depends on the affective constitution of agents and/or appraisers, we must always ask, if benevolence grounds reasons for acting morally, what kind of benevolence do human beings have by their very frame? It is clear that Hutcheson maintains that, while human agents have both motives, the particular benevolent motives are naturally stronger. It is only by intentionally undergoing a process "enlarging and abstracting" that there arises within human agents the calm universal benevolent motives.[60] What I need to show now is that, given the typical strength of particular benevolence over universal benevolence, it may turn out to be the case that a benevolence-grounded reason does not provide an agent with a reason to be moral.

In order to establish the second part of my thesis, I need to specify a feature of my discussion that up to this point has been rather vague, *viz.* what it means on Hutcheson's account to *act morally*. All I want to say about what it means to act morally on Hutcheson's view is that, first, whatever it is, it will involve acting from benevolent motives and that, secondly, Hutcheson's view allows that moral actions can be arranged on a continuum from better to worse: some actions are morally better than others, some are worse, and some are the very best of moral actions available. He says for example, "we may see what Actions our moral Sense would most recommend to our Election, as *the most perfectly Virtuous*: viz. such as appear to have the most universal unlimited Tendency to the greatest and most extensive Happiness of all the rational Agents, to whom our Influence can reach" (*Inquiry* III.x; emphasis added). Hutcheson is clear that all benevolence is approved of by the moral sense, but he adds that we more strongly approve of those actions that are motivated by calm universal benevolence: "Our

---

[60] This is not necessarily the case with respect to benevolent agents. Hutcheson maintains that there could be beings that are naturally possessed of the calm universal benevolence alone (angels, for instance), but he denies that this is true of human beings.

moral Sense, tho it approves all particular kind Affection or Passion, as well as calm particular Benevolence abstractly considered; yet it also approves the Restraint or Limitation of all particular Affections or Passions, by the calm universal Benevolence" (EPA II.ii, R 357).[61] The best actions, or perhaps those that are morally right, are those that bring about the greatest happiness for the greatest number. This does not imply that actions motivated by benevolence towards some more limited sphere are not morally praiseworthy. However, the approval of partial benevolence depends on the actions to which this motive gives rise *not being inconsistent with a greater good for a greater number*. When there is such an inconsistency, the moral sense will not approve of the more limited benevolence at all, but will continue to approve of universal benevolence. However, as we saw above, when there is such a conflict between particular and universal benevolence the motive that wins out will simply be whichever one is stronger. What this suggests is that, while an agent has some reason to perform a benevolent act, her moral sense approves of more strongly her performing an act from universal benevolence. This, however, does not yet give her a reason to act from universal benevolence: her affective constitution is not directed towards the good of the greatest number but is instead directed at some more limited system (such as her family, her neighborhood, a faction, or sect). So, what is needed here is some *further* exciting reason. The reason that is missing is not a reason to pursue other-directed interests over self-interest, it is a reason to pursue the greatest happiness of all concerned over the happiness of some more local sphere.

I think that Hutcheson does have a response to this question, and it is one of the responses that Radcliffe identifies. However, I want to make it clear that I am using this response to answer a different question than it seems Radcliffe is asking. What this will amount to, I hope to show, is that Radcliffe's presentation of these two possible responses on behalf of Hutcheson to the "why be moral" question work together. Nothing Radcliffe says rules out this possibility. But the

---

[61] The passage quoted is incomplete. Hutcheson goes on to add that such restrictions of particular benevolence by universal benevolence are the only way to secure constant self-approbation. I will have more to say about this below.

account of how they work together is also not developed. What I aim to do now is to specify the connection between them.

First, we must recall the other account of the reasons that an agent has for acting morally that Radcliffe identifies. This is the account that holds that an agent has a reason to be good in virtue of the fact that reflective self-approbation triggered by her benevolently motivated acts will give rise in her to pleasure. Acting morally in this way indirectly results in the agent acquiring a greater amount of natural goodness. Hutcheson says, "To make this Desire [calm universal benevolence] prevalent above all particular Affections, is the only sure way to obtain constant Self-Approbation" (EPA II.ii, R 357). In addition he says, "if we must bring in Self-Love to make Virtue Rational, a little Reflection will discover…that this Benevolence is our greatest Happiness; and thence we may resolve to cultivate, as much as possible, this sweet Disposition, and to despise every opposite Interest" (*Inquiry* III.xv). It is important to keep in mind that a motive from self-love to obtain the pleasant sensation of reflective self-approbation would be entirely undermined were it the agent's only motive. We cannot be virtuous from this motive alone, and if the pleasure of reflective self-approbation attends only our perception of ourselves as having acted virtuously, then we must suppose that our action was at least in part motivated by a genuine love for others. The idea here, then, is that self-love may provide me with reason to strengthen my benevolence, or rather to widen its scope. Hutcheson says, "Let the Misery of excessive Selfishness, and all its Passions, be but once explain'd, that so Self-love may cease to counteract our natural Propensity to Benevolence, and when this noble Disposition gets loose from these Bonds of Ignorance, and false Views of Interest, *it shall be assisted even by Self-love*, and grow strong enough to make a noble virtuous Character" (*Inquiry* VII.ii; emphasis added). That Hutcheson is here referring to a conflict between self-love and benevolence does not diminish the importance of his claim for my suggestion. He clearly thinks that self-love is an important component in the development of a virtuous character. My suggestion, building off Radcliffe's, is that self-love provides us with an exciting reason to expand our sphere of concern.

Thus, we can see the two accounts provided by Radcliffe working together in an interesting way. I have a reason to be moral insofar as acting morally actually satisfies the interests that I currently have. That is, my having a reason to be moral depends on the fact that I already care about the well-being of others for its own sake. However, since most of my concern for others is focused on a limited sphere of others (and while this may usually be in fact conducive to the overall good) there is no reason to think that I will have a reason to pursue the morally best course of action: I simply do not have a stronger desire for universal good than I do for more localized goods. However, I do have a reason to strengthen my benevolent motives insofar as doing so will result in greater happiness for me. I receive pleasure when I act from motives of benevolence, and the greater the extent of the benevolent concern, the greater the pleasure I take in being so moved.

Concluding Remarks

Hutcheson's views on obligation and motivation are very interesting in their historical context. As we have seen, he rejects the prevailing view that moral obligation is to be understood in terms of a rational motive. Moreover, it is hard to see any strict sense in which Hutcheson accepts any version of internalism. Nevertheless, he does want to maintain a very strong, if contingent, relationship between obligation and motivation. Given our psychological makeup, if we approve of our doing some action, and disapprove our omitting it, then we must recognize that we have a particular sort of desire, *viz*. a desire for the well-being of others for its own sake. Our states of approval and disapproval always track the motivational states of ourselves or others. This feature of Hutcheson's view seems to explain, in part, why his view is difficult to classify as a version of externalism. We also saw that Hutcheson does not accept that moral motivation involves a desire for moral goodness: all motivational states that have some good as object (which again is not every motivational state) take *natural* good as object.

According to Hutcheson, we ought to pursue virtue, in the interested sense of obligation, because doing so will yield a greater amount of happiness for ourselves. But why *morally* ought

we to do what we morally ought to do? That is, what sort of *disinterested* reason do we have for doing what we morally ought? Here, Hutcheson only has recourse to contingent features of human psychology. The only sort of disinterested reason we could have comes to us from the desires that are contingently related to our possessing a public sense. In other words, all Hutcheson can say on this point is that we ought to do what we morally should because we have desires for the well-being of others. Hutcheson does not seem to think that there are any human beings who lack these desires entirely. He recognizes that the public affections are often obscured to us by the operation of more limited affections and by the passions. So, he accepts that everyone can come to have calm universal benevolence, so long as he uses his reason to clear away its obstruction by other affections and passions.

Now, we might ask Hutcheson "why ought I, in the disinterested sense, to remove the obfuscating influence of the passions?" Besides that he thinks it would make us happier to be governed by the calm affection, than by either the passions or the passionate desires, what I suspect Hutcheson would say here is that to do so would simply be the natural outgrowth of our continuing this line of questioning. What I mean is that asking for a reason, not rooted in self-interest, why we ought to inculcate in ourselves a calm universal benevolence presumes that we already accept that reason should, in some way, guide our conduct. If we continue to use reason *as a guide to our conduct* then, Hutcheson thinks, reason will disclose to us the various objects and actions in the world that will create pleasure for ourselves or others. We will come to see that the well-being of the most extensive system of sentient creatures results in the greatest amount of pleasure. Having apprehended this good, we will come to desire it. Using reason in the way that we already presume it can, or should, be used will inevitably lead us to have calm universal benevolence. The amoralist, then, is not someone who is presenting a well-founded challenge to moral action. Rather, the amoralist is someone who, presuming they are a normal human agent, has not yet completed his line of inquiry: he has not used his reason sufficiently and so has yet to acquire calm benevolent motives.

This, I think, is one of the reasons why Hutcheson takes his moral sense theory to be advantageous over egoism and rationalism when it comes to the question of the practicality of ethics. Not only does he have independent reasons for rejecting both egoism and rationalism, he also thinks, as we saw, that such "schemes" fail to inculcate a virtuous character. Hutcheson does not think there is any necessary connection between ones coming to recognize certain truths and ones forming certain desires or dispositions. However, he does think that there is a very strong contingent connection between the apprehension of some truths and the formation of certain desires. The above discussion should make this clear: by coming to understand that certain things are good, we tend to form desires for them. Hutcheson makes sure to argue that the highest good that can be achieved is the pleasure of the greatest number. By recognizing this, my constitution is such that I will come to desire it for its own sake. So, there is nothing in particular about *Hutcheson's system* that provides this advantage: what is advantageous about it is only that it goes to great length to champion the pleasure of the greatest number. Hutcheson says,

> Now the principal Business of the moral Philosopher is to shew, from solid Reasons, "That universal Benevolence tends to the Happiness of the Benevolent, either from the Pleasures of Reflection, Honour, natural Tendency to engage the good Offices of Men, upon whose Aid we must depend for our Happiness in this World; or from the Sanctions of divine Laws discover'd to us by the Constitution of the Universe;" that so no apparent Views of Interest may counteract this natural Inclination: but not to attempt proving, "That Prospects of our own Advantage of any kind, can raise in us real Love to others. (*Inquiry* VII.ii)

It is because of his attempts to accurately characterize the relationship between universal benevolence and self-interest, that Hutcheson thinks he is able to predict that his moral philosophy will have a greater practical effect than that of his contemporaries.

CHAPTER 5

HUME'S THEORY OF THE PASSIONS

Introduction

There is no doubt that Hume owes much to Hutcheson. One commentator dramatically illustrates this debt by musing that Hume wrote Book 3 *A Treatise of Human Nature,* "Of Morals," with a copy of Hutcheson's *Essays and Illustrations* on his desk for ease of reference.[1] However, there is little doubt that Hume found much in Hutcheson with which he disagreed. There are very clear differences between their respective accounts of the passions. Given the centrality that the passions play in both systems, understanding Hume's theory is integral both to understanding his moral philosophy as a whole, and as it compares to Hutcheson's. In this chapter, I present some reasons for thinking that an understanding of Hume's theory of the passions is important for understanding his moral theory. I then present Hume's theory of the passions, at least those elements of the theory that relate to his theory of moral motivation and moral obligation which include the motivational limitations of reason alone and the nonrepresentational nature of the passions. I conclude with a discussion of the nature of the will and necessity as a way to enter into Hume's account of moral evaluation. As with Hutcheson, there is controversy surrounding how to understand Hume's metaethical commitments, and I will not here attempt to settle this dispute. However, I will provide a brief discussion of the problem.

Book 3 in Relation to Earlier Parts of the *Treatise*

Book 3 of the *Treatise* was published a few years after Books 1 and 2, "Of the Understanding" and "Of the Passions" respectively. This fact, together with Hume's claim that this third book of the *Treatise* is "in some measure independent of the other two, and requires not that the reader shou'd enter into all the abstract reasonings contain'd in them," might lead one to suspect that the conclusions at which Hume arrived in the previous books are simply not relevant

---

[1] Citation required.

to his account of moral philosophy (T *Advertisement*, 292).[2] Moreover, given a skeptical interpretation of the conclusions of Book 1, it might become even more difficult to think that all books of the *Treatise* could be unified in any serious way.

Even if one were to hold this sort of a view about the *Treatise* as a whole, it is undeniable that Books 2 and 3 present mutually dependent and reinforcing claims. One cannot understand Hume's moral theory without understanding his moral psychology, at least those parts of it that relate to his account of the passions, motivation, and the influence of reason.[3] Moreover, at the very beginning of Book 3 Hume says, "the present system of philosophy will acquire new force as it advances," signaling that the discussion of Book 3 should not be thought to be *wholly* independent from the previous sections of the *Treatise* (T 3.1.1.1, SBN 455). Rather, "Of Morals" should, Hume hopes, "corroborate whatever has been said concerning the *understanding* and the *passions*" (ibid.).

It is not my intention here to settle any disputes about whether Hume is best understood as a skeptic nor, if he is, what his brand of skepticism looks like. I assume that Hume had reached an epistemological starting point that he took to be adequate for addressing the questions of human psychology and morality, but I will not here evaluate the plausibility of this starting point.[4] In what follows I will assume only that Hume's moral theory requires an understanding of his theory of the passions while being simultaneously an extension and application of that theory. This is not to say that all of the discussions from Book 1 are irrelevant; Hume's theory of

---

[2] With the exception of references to the Preface and the Advertisement, I follow what is fairly standard practice in citing Hume's *Treatise*. I use the Norton edition (Oxford: 2000) and all references to it will be abbreviated with a T followed by book number, part number, section number, and paragraph number. I include references to the Selby-Bigge edition revised by Nidditch (1978) abbreviated 'SBN' and followed by page numbers as indicated in the Norton.

[3] This is a weaker version of the interpretive claim Baier uses in *A Progress of Sentiments*, which maintains that all the books of the *Treatise* are mutually dependent and reinforcing. See Baier 1991. Loeb 1977 denies that the connections between Books II and III are as strong as I am here making them out to be.

[4] For a non-skeptical reading of the *Treatise* see Baier 1991.

the passions draws heavily from his account of personal identity and the perception of an enduring self,[5] and is itself an extension of the psychological principles of association he develops to account for our causal beliefs and our beliefs in continued unperceived existence. And while there are places in Hume's discussion of psychology and morality where the arguments regarding our ideas of cause and effect and of necessary connection are important to recall, it will not be necessary to settle any major interpretive debates about these arguments.

That Hume's discussion of the psychology of the passions and his discussion of morality are interrelated is obvious from the very start of Book 3: "If morality had naturally no influence on human passions and actions, 'twere in vain to take such pains to inculcate it" (T 3.1.1.5, SBN 457). Two paragraphs later Hume recalls his argument from Book 2 "that reason is perfectly inert" (T 3.1.1.8, SBN 458; T 2.3.3, SBN 413-419). These are the crucial premises in what is perhaps Hume's most famous argument for the claim that, contrary to Clarke, Cumberland, and Wollaston, moral distinctions are not derived from reason. The relationship between human psychology and morality allows Hume to reject the rationalist's view that "morality, like truth, is discern'd merely by ideas, and by their juxta-position and comparison" (T 3.1.1.4, SBN 456). Hume claims that human action is deeply rooted in the passions and his claim that morality affects passions and actions in a way that reason alone is incapable of shows us that there is a close tie between Hume's psychology of the passions and his moral psychology.[6]

Before turning to a more detailed examination of Hume's argument here, it is worth recalling the methodological precepts underlying the "science of human nature" as it's developed throughout Hume's corpus. Hume maintains that the only objects that are immediately before the

---

[5] Penelhum 2009. Baier 1991 discusses the relationship between the passions and the account of the self in Ch. 6.

[6] I do not mean for this to be a satisfactory presentation of Hume's Motivation Argument at the beginning of T 3.1.1. I am here only trying to show one way in which Hume's moral theory depends on his theory of moral psychology: this dependence is clearly illustrated in this argument, even when the argument is presented in the rather loose way that I have done. I will return to consider the Motivation Argument in the next chapter.

mind are perceptions: "nothing is ever really present with the mind but its perceptions or impressions and ideas…To hate, to love, to think, to feel, to see; all this is nothing but to perceive" (T 1.2.6.7, SBN 67). Hume claims that all perceptions are either impressions or ideas. Impressions are distinguished from ideas by their "force and liveliness" and so they "enter [the mind] with most force and violence" (T 1.1.1.1, SBN 1). Hume holds that sensations, passions, and emotions are all impressions. Ideas, on the other hand, are "the faint images of [impressions] in thinking and reasoning" (ibid.).[7]

After noticing this general difference between the two kinds of perception, a difference that he labels as the difference between "feeling and thinking," Hume further claims that, "the first circumstance, that strikes my eye, is the great resemblance betwixt our impressions and ideas in every other particular, except their degree of force and vivacity" (T 1.1.1.3, SBN 3).[8] Given the "constant conjunction of resembling perceptions," Hume argues that simple ideas causally depend on the existence of simple impressions (T 1.1.1.8, SBN 4). This is Hume's copy principle: "that all our simple ideas in their first appearance are deriv'd from simple impressions, which are corresponding to them, and which they exactly represent" (T 1.1.1.7, SBN 4). The copy principle consists, then, of two distinct theses: *the resemblance thesis* – ideas and impressions resemble each other in content, but not in force or vivacity; *the dependence thesis* – simple ideas depend for their existence on simple impressions. The copy principle forms the bedrock of Hume's empiricist methodology because he requires that all ideas, especially those complex ideas purportedly referred to in typical philosophical discourse like *cause*, *necessary*

---

[7] Impressions and ideas are not different in kind, but are different in their degree of force and liveliness. There are other differences that Hume notes between impressions and ideas. For example, he claims that impressions can be "blended so perfectly together" and have a "total union" with each other. Ideas, on the other hand, have "a kind of impenetrability, by which they exclude each other" and can be compounded but not mixed together (T 2.2.6.1, SBN 366). Kendrick 2009 argues that what is important about the relationship between ideas and impressions is that ideas possess the same experiential content as the impressions which cause them.

[8] Hume thinks the resemblance is strongest between our simple impressions and our simple ideas; complex ideas need not resemble complex impressions, although complex ideas, ultimately, must have simple ideas as parts that themselves resemble some impression or other.

connection, *self*, and *virtue*, be traced to the simple impressions from which they, ultimately, arise.[9]

With this as his methodological requirement, Hume turns in Book 3 to ask, "Whether 'tis by means of our ideas or impressions we distinguish betwixt vice and virtue, and pronounce an action blameable or praise-worthy?" (T 3.1.1.3, SBN 456). After asking this question, Hume immediately launches into his criticism of rationalism. This suggests that the main issue between Hume and his rationalist opponents turns, as perhaps is to be expected, on a question of moral epistemology. The conclusion of the argument in 3.1 is that "The rules of morality…are not conclusions of our reason" (T 3.1.1.6, SBN 457). It is important not to overstate this conclusion; reason does play a role in our lives as moral agents, and I will return later to consider what Hume takes that role to be. The conclusion of this particular argument is that the rationalists are wrong in thinking that moral distinctions can be discerned in our ideas and their relations by the use of reason *alone*.

Since Hume, like Hutcheson, takes the rationalists to maintain that "virtue is nothing but a conformity to reason," and that "there are eternal fitnesses and unfitnesses of things, which are the same to every rational being that considers them," he denies that any operation of reason is sufficient to perform the tasks it would need to if it were the sole epistemological faculty underlying the apprehension of virtue and vice (T 3.1.1.4, SBN 456). In line with Shaftesbury and Hutcheson, Hume maintains that morality is dynamic and practical. Our perceptions of virtue and vice are not indifferent to us. Perceiving that an action is done from maliciousness or contempt does not leave our passions unaffected: we *hate* the vicious character performing such acts. With respect to the agent acting from virtuous motives, on the other hand, we find that such a person inspires our *love*, our *esteem*, and our *trust.* The passionate states that always come

---

[9] The case is, of course, complicated by Hume's discussion of the missing shade of blue. There has been much written about the significance of this purported counterexample to the dependence thesis. Since Hume clearly takes himself to have established the methodological viability of the copy principle by Book III, and since the missing shade of blue is present also in the first *Enquiry*, it seems fairly safe to grant him the principle in "Of Morals" and in the second *Enquiry*.

along with our moral perceptions likewise have an influence on our actions. Knowing the virtuous thing to do does not leave me indifferent to its pursuit; knowing the vicious, likewise does not leave me indifferent to its avoidance. In Chapter 6 we will see that there are two importantly different ways to understand Hume's commitments here regarding the practicality of ethics. What is important to note here is Hume's commitment to the claim that there is some connection between our apprehension of moral distinctions and both our passions and our actions.

The next element in Hume's argument against the rationalists is his claim regarding the practical impotence of reason alone. It depends on Hume's famous doctrine that "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (2.3.3.4, SBN 415). In order to understand this claim, it will be helpful to have a greater understanding of Hume's account of the passions in general and his view of the nature of reason.

Hume's Theory of the Passions

I cannot hope to fully do justice to the richness and complexity of Hume's theory of the passions in this section, and so set myself far smaller goals. I aim to show how Hume's theory of the passions fits within his empiricist methodical constraints, how Hume conceives of the content of the passions, and, ultimately, how these views connect up with Hume's account of reason, morality, and action. However, these smaller aims meet with some fairly large interpretive challenges. I will do my best to signal the interpretive difficulties as they arise, and will occasionally suggest some rough interpretive sketches to address them, but I do not pretend that what I have to say in this section could stand on its own as a contribution to many questions about Hume's theory of the passions. Making such a contribution is not my purpose here.

The Multiple Divisions of the Passions

Locating the passions within Hume's empiricist methodological constraints is a rather straightforward task, though not one that can be done without comment. Early on in Book 1,

Hume divides the impressions into the impressions of sensation and the impressions of reflection. In Book 2, he returns to the same distinction and this time he additionally calls them original and secondary impressions.[10] The passions are kinds of secondary impressions, or kinds of impressions of reflection. Unlike the impressions of sensation, which arise "in the soul originally, from unknown causes" that we ordinarily attribute to the existence and operation of external objects, the impressions of reflection arise through entirely internal *mental* processes.[11] We have impressions of sensation when an "impression first strikes upon the senses" and then is copied as an idea that resembles it in all respects except force and vivacity (T 1.1.2.1, SBN 8). One particularly important pair of impressions that subsequently get copied as ideas are the sensations of pleasure and pain. When the idea of pleasure or pain is again presented to the mind, either directly through memory or through the association of present sensations with remembered ideas, the mind responds by producing a very forceful and lively perception: "The idea of pleasure or pain when it returns upon the soul, produces the new impressions of desire and aversion, hope and fear, which may properly be call'd impressions of reflection because deriv'd from it" (ibid.).

Hume does not intend for the forgoing to be an exhaustive account of the origination of all of the passions, but means instead to signal that the passions are *impressions* that arise only through the operation of other mental phenomena rather than merely through the operation of the sense faculties. The impressions of reflection cannot arise independently from other perceptions, unlike the impressions of sensation.[12] This is true even if there are important differences in the

---

[10] David Owen argues that Hume *replaces* the Book 1 distinction between impressions of sensation and impressions of reflection for the more precise Book 2 distinction between original and secondary impressions. See Owen 2009, 76-7.

[11] As Norton points out, Hume "makes no use of speculative physiology" to explain the origin and operation of the passions (2009, I47). This marks a departure from Hobbes's account of the passions (desire, aversion, joy, fear, etc.), in which the passions are understood entirely in terms of physiological processes (helping/hindering the vital motion around the heart, for instance).

[12] Norton 2009, 490.

particular causal processes by which some passions arise. For instance, Hume distinguishes between the *direct* and the *indirect* passions in terms of a difference between the casual processes by which they arise. He says that the direct passions "arise immediately from good or evil, from pain or pleasure" (T 2.1.1.4, SBN 276). The direct passions include desire and aversion, grief and joy, and hope and fear. These are all impressions even though there are occasions on which they arise without proceeding from the copied sensations of pleasure and pain. Hume says that the direct passions arise immediately from the perception of pleasure and pain: desire and aversion arise from my sensations of pleasure and pain and also from my remembering that pleasure and pain are likely to attend other perceptions with which they have been previously conjoined.

Before looking briefly at Hume's account of the indirect passions, it is important to note a complication to the foregoing description of the direct passions. Hume adds that "the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable" and he gives the following as examples: "the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites" (T 2.3.9.8, SBN 439).[13] These direct passions are related to the calm passions to which Hume had previously referred: "Now 'tis certain, there are certain calm desires and tendencies, which, tho' they be real passions, produce little emotion in the mind" (T 2.3.3.8, SBN 417). The examples he provides here include "instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children," as well as "the general appetite to good, and aversion to evil, consider'd merely as such" (ibid.). These passions are what Norton refers to as *productive* as opposed to *responsive* passions.[14] The passions may cause an agent to experience pleasure or pain, but they do not exist for the sake of experiencing such pleasure or avoiding

---

[13] I here follow Penelhum's 2009 discussion of this species of direct passions.

[14] Norton 2009, I48. See chapter 6 below for further discussion of these passions.

such pain.[15] They are simply fundamental and original operations of human psychology. This class of passions is very important for Hume's moral theory and his account of moral motivation, and I will return to discuss what makes these passions "calm" in what follows.

To return to the direct/indirect distinction, Hume initially describes the causal process by which the indirect passions arise rather vaguely: "By indirect [passions I understand] such as proceed from the same principles [good or evil, pleasure or pain], but by the conjunction of other qualities" (T 2.1.2.4, SBN 276). Very briefly, I may feel pride towards one of my beautiful possessions. The pleasure I take in the object is an impression of sensation, and this may have initially spurred me to acquire the object by giving rise in me to a desire for it (a direct passion). However, once it is mine and it still serves as a source of pleasure, I take pride in it in virtue of its being something *of mine*.[16] So this passion arises, in part, through the production of pleasure or pain but also requires the operation of principles of association to incorporate the idea of self.[17] It should not be thought that the indirect passion's arising in this more complicated causal process means that it is not an impression.[18] Recall that what is distinctive about impressions is their forcefulness and liveliness. Hume insists that the principles of association of ideas (the natural relations of resemblance, contiguity, and causation) together with the analogous principle

---

[15] I think that these remarks are sufficient to rule out any *straightforward* identification of Hume as a psychological hedonist.

[16] As Hume would say, I am the object of pride (and humility), and the beautiful object is the cause of my pride. Every cause of pride or humility is divided into a quality and its subject. The beautiful object could never be a cause of humility since the quality, beauty in this case, is productive of pleasure which is not associated with the emotion of humility. However, were the same subject to deteriorate from beauty to deformity, then it would serve as a cause of humility for me.

[17] Of course there are details to this process that I have left out for the sake of brevity. One crucial component of Hume's theory is the "double relation of ideas and impressions" that gives rise to particular passions. I mention this omission because the "double relation" arises out of Hume's use of the principles of association developed with respect to ideas in Book I to account for the origin and operation of the passions. His use of these principles provides a clear illustration of the continuity present between all three books of the *Treatise*: love is a passion that arises from the principles of association and is, as we will see, importantly related to moral approbation.

[18] Moreover, Hume says that we cannot provide an account of the passions because they are simple, and so are indefinable. We can account for their origin and operation even though they are simple and so we cannot give an account of them.

of association of impressions (resemblance) facilitate the movement between various perceptions. This movement is capable of transferring the forcefulness and liveliness of an impression to an idea in such a way that the resultant perception takes on the force and vivacity of its causal antecedent.[19] This transference of force and vivacity according to the principles of association allows Hume to have a rather complicated causal account of the indirect passions while still maintaining that they are simple impressions of reflection.

It will be helpful to say just a bit more about the causal mechanism that Hume thinks is involved in the production of the indirect passions. This mechanism is the "double relation of ideas and impressions" and Hume's exposition of its operation is perhaps the most elaborate and technically sophisticated element of his science of human nature. I aim here only to provide a very rough sketch of it.[20] Let's continue with the example of my taking pride in one of my beautiful possessions (though Hume is clear that I can take pride in many different things: my virtue, the success of my family, my bodily strength, and etc.). Hume distinguishes the causes of the indirect passions from their objects. The causes of indirect passions are then further distinguished into subject and qualities. The subject is that of which I have an idea, namely my possession. My idea of the subject calls to mind via the principles of association my idea of myself since it is, after all, an idea of *my* possession. Recall that Hume thinks that pleasure and pain are partly causally responsible for the production of the indirect passions. The quality of the subject is what produces in me sensations of pleasure or pain: in our example it is the beauty of my passion that gives rise in me to pleasure. Here Hume adds to his associationist psychology the thesis that there are associations of resemblance between impressions. The pleasure I experience from my beautiful possession resembles the pleasurable passion of pride. This in turn causes me to feel pride, and since the subject and quality have also brought to mind my idea of

---

[19] "I wou'd willingly establish it as a general maxim in the science of human nature, that when any impression becomes present to us, it not only transports the mind to such ideas as are related to it, but likewise communicates to them a share of its force and vivacity" (T 1.3.8.2, SBN 98).

[20] I closely follow Norton's description of this mechanism. See Norton 2009, I51.

myself, my feeling of pride has myself as its object. Hume maintains that the object of pride is always oneself since we take no pride in what is not related to us at all.

Hume says that all passions are simple impression so it should not be thought that he is intending to provide an analysis of the indirect passions. Nancy Schauber raises a problem with Hume's account here. She says that Hume holds that every passion has its own "distinctive felt quality, and that each passion can be identified by its distinctive feeling."[21] But she thinks that Hume is inconsistent because he both holds that each passion can only be identified by its distinctive feeling, and also that each passion has to be identified by the causal mechanisms that produce it. She doubts that Hume's psychological mechanisms will provide us with an adequate way to correctly identify and distinguish different passions. However, it is not clear that Hume is committed to the view that passions can only be identified by their distinctive feelings. Schauber provides no textual evidence for this claim. As I argue in the next chapter, Hume can maintain that all passions have a distinct feeling but people inexperienced at introspection may find it difficult to be able to tell the difference between them since many of these feelings will be rather similar. I am assuming here that Hume's account of the double relation of ideas and impressions is an account of the causal origin of passions with different felt qualities.[22]

Schauber identifies a further difficulty with Hume's account of the passions. This difficulty arises from Hume's use of the term 'emotion' in relation to the passions. The term 'passion' has fallen out of general philosophical use and where Hume talks about passions, philosopher's today would likely prefer to talk about emotions. But, as Schauber points out, it is less than clear exactly what Hume's account of emotion is: "He sometimes seems to mean 'a conscious mental state which accompanies, but is not identical to, a passion.' At other times, Hume writes as though emotions are simply low-grade, washed-out passions."[23] Hume also

---

[21] Schauber 1999, 345.

[22] I follow Penelhum in this account. See Penelhum 2015, 208.

[23] Ibid., 363.

says, as we'll see below, that pride and humility are "pure emotions in the soul" (T 2.2.6.3, SBN 367). This latter sort of passage leads me to think that it is fairly safe to identify passions with emotions, but I do not think that the issue is perfectly clear in Hume.[24]

Calm and Violent Passions

Hume's distinction between the direct and the indirect passions is his own contribution to the modern understanding of the passions.[25] He also appeals to a set of distinctions that are more familiar: He distinguishes on the one hand calm from violent passions,[26] and on the other hand he distinguishes strong from weak passions. I will have more to say about the latter distinction in what follows, but for now it will be helpful to initially characterize the strong/weak distinction in terms of the motivational force of the passion. Introducing the former distinction, Hume says:

> Of the first kind [calm passions] is the sense of beauty and deformity in action, composition, and external objects. Of the second [violent passions] are the passions of love and hatred, grief and joy, pride and humility. This division is far from being exact. The raptures of poetry and music frequently rise to the greatest height; while those other impressions, properly call'd *passions*, may decay into so soft an emotion, as to become, in a manner, imperceptible. (T 2.1.1.3, SBN 276)

Here we encounter an interpretive challenge. Hume insists that the passions are impressions. But we might well wonder how this could possibly be when he allows for the existence of calm passions. Hume recognizes that some emotional responses "may decay into so soft an emotion" that they become "in a manner, imperceptible." Impressions, as perceptions, seem hardly the sorts of things that could become imperceptible in any sense.[27] It doesn't seem

---

[24] I think the identification is fairly safe and in so thinking follow Penelhum who often replaces talking about passions with talking about emotions. See Penelhum 2015, 214.

[25] Jane McIntyre argues that both the terminology and the division of the passions into direct and indirect were entirely original to Hume. See Macintyre 2000, 78.

[26] This distinction is related to Hutcheson's distinction between the calm and violent affections, but for him, unlike for Hume, the *passions* are always violent so that the notion of a calm passion is incoherent to Hutcheson.

[27] Penelhum 2009, 249; 2015, 209. Paxman 2015 offers a novel interpretation of the calm/violent distinction in large part to overcome this problem, which she dubs the "imperceptible perceptions" puzzle.

we can pass this off as an instance of rather careless wording on Hume's part. There seems to be something fundamentally problematic with the very notion of a *calm* impression given that Hume introduces impressions in Book 1 as those of our perceptions that are the more forceful, lively, and *violent* (T 1.1.1.1, SBN 1).

The standard way of understanding Hume's distinction between calm and violent passions is to understand it as describing different phenomenological feels of the passions. I understand the idea to be similar to Hutcheson's talk of the bodily perturbations characteristic of the passions as he conceives them. For Hume, the violent passions are the ones that strike upon the soul in such a way that they produce rather noticeable and forceful perturbations.[28] We usually cannot help but notice when we're overcome with love, joy, or pride and it is this emotional feeling that Hume is signaling when he describes passions as being violent. Those passions that produce a sensible agitation when they arise are violent, and those that arise without such an agitation are calm.[29] It is not clear whether Hume intends for the calm/violent distinction to cut across all categories of passions,[30] or if Hume intends for the calm/violent distinction to be fundamental and the direct/indirect distinction to only further divide the violent

---

Since I do not find this to be as much of a puzzle as Penelhum and Paxman have made it out to be, I will not discuss Paxman's novel interpretation of the calm/violent distinction.

[28] This would make Hume's account of the violent passions rather similar to Hutcheson's account of the passions *simpliciter*. For the time being, it will be helpful to note that there is at least one crucial difference between them. Hutcheson insists that the bodily perturbations of the passions interfere with the operations of reason. As we will see, Hume does not think this is usually an interference with *reason* as such, and certainly does not think that such an interference is essential to the nature of violent passions.

[29] Paxman denies that a violent passion is merely a passion attended with sensible agitation (2015). She maintains that any state of mind is violent when it interrupts or disrupts the regular natural or customary movements between perceptions. However, she does recognize that different sorts of phenomenological feeling accompany calm and violent passions. For this reason, it is not clear that she has offered a new interpretation of the calm/violent distinction, or has instead offered a very compelling account of the causal mechanisms that give rise to this difference in sensible agitation.

[30] Kemp Smith (1941), Árdall (1966), Penelhum (2009), and Paxman (2015) accept this view.

passions.[31] I think there is good reason to suppose that Hume did not intend for particular kinds of passions to be always calm or always violent.[32] However, it is likely that he thought that certain kinds of passions are *generally* calm and that some others are *initially* violent. There is no a priori reason why every passion might not be calm,[33] and Hume does seem to allow that there are in fact psychological mechanisms in place by which many (if not all) passions could become calm: "when a passion has once become a settled principle of action, and is the predominate inclination of the soul, it commonly produces no longer any sensible agitation. As repeated custom and its own force have made everything yield to it, it directs the actions and conduct without that opposition and emotion, which so naturally attend every momentary gust of passion" (T 2.3.4.1, SBN 419). So long as a passion is capable of becoming a "settled principle of action" then once it has become such it will be felt calmly.

Radcliffe maintains that the violence of the passions is the same feature as the force and vivacity of the impressions of sensation referred to in Book 1. She argues that Hume's psychological theory would be unduly complicated were Hume to introduce violence as a unique feature of mental events that is so similar to, but nevertheless distinct from, force and vivacity. She also claims that understanding violence as force and vivacity helps to explain why it is we mistake, as Hume argues that we do, the deliverances of calm passions for those of reason: insofar as reason deals with ideas, or perceptions that are very low on the force and vivacity continuum, they can easily be confused with other perceptions that are low on this same continuum, namely the calm passions. This seems to exacerbate the imperceptible perception

---

[31] Loeb accepts this view (1977). Radcliffe says that Loeb's taxonomy is "probably the best" (2015, 551). However, she finds a great deal at issue with his account and develops her own taxonomy that seems in many ways more similar to Norton's taxonomy than it does to Loeb's.

[32] As we have seen at T 2.1.1.3, this is because Hume calls the calm/violent distinction "vulgar" and "specious."

[33] Árdall (1966) and Loeb (1977) agree about this much at least, so the question then should not be understood as s conceptual question concerning the passions, but should instead be understood as a question of psychological matter of fact.

puzzle.[34] If a calm passion is capable of being confused with an idea, then in what sense is it an impression at all? Its lacking in force and vivacity, which Radcliffe claims is the same as its lacking in violence, seems to take away exactly that feature that is distinctive of impressions.[35]

I would like to propose a rough response to this purported puzzle. First, it is worth recognizing that if there is a problem with calm passions in virtue of the fact that they are impressions purportedly nearly devoid of force and vivacity, this is a problem for Hume's account of impressions in general. The claim that impressions may be rather lacking in their characteristic force and vivacity is a phenomenon to which Hume refers in his introduction of the distinction between impressions and ideas: "it is not impossible but in particular instances [both kinds of perceptions] may very nearly approach each other. Thus in sleep, in a fever, in madness, or in any very violent emotions of the soul, our ideas may approach to our impressions: As on the other hand it sometimes happens, *that our impressions are so faint and low, that we cannot distinguish them from our ideas*" (T 1.1.1.1, SBN 2, emphasis added). So, *if* there is a puzzle here about the calm passions, it is a puzzle that arises for Hume's theory of impressions as such.

Now, it is important to see that Hume only says that the calm passions are *in a manner* imperceptible. He later says that the calm passions "produce little emotion in the soul, and are more known by their effects than by the immediate feeling or sensation" (T 2.3.3.8, SBN 417). It is interesting that Hume characterizes the operation of the principles of association in a similar way: "it is evident, that the association of ideas operates in so silent and imperceptible a manner, that we are scarce sensible of it, and discover it more by its effects than by any immediate feeling or perception" (T 2.1.9.4, SBN 305). Of course, Hume does not think that the association of ideas is an impression. What is more, he does not think that the associative mechanisms must

---

[34] This is Paxman's name for the problem identified by Penelhum (2015).

[35] To add to the complications here, Hume also maintains that ideas can be converted into impressions via associationist mechanisms that transfer forcefulness and vivacity. This is, roughly, the way in which sympathy operates in the *Treatise*. Now the question becomes, how can there be forceful and lively *ideas* as well as ideas that *become* impressions in virtue of their being forceful and lively? See Ch. 6 for more on the operation of sympathy in the *Treatise*.

always evade our attention. We can discover these principles of association even though they are the sorts of things that would likely evade our direct introspective awareness. Hume is saying that we have a better shot at apprehending the associations between ideas if we attend to their effects than we would if we were to merely try carefully to attend to them. The same goes for the calm passions: they are not perceptions that we paradoxically cannot perceive, but rather are perceptions that are better grasped via their causal effects than through careful introspection on the perception itself. They are not *unfelt* perceptions. They are, instead, perceptions that are *barely* felt and so are best discovered indirectly by attending to their causal effects. I cannot here provide a more detailed response to this worry.

Weak and Strong Passions

Above we saw that Hume holds that a passion can become calm once it is a "settled principle of action" (T 2.3.4.1, SBN 419). Becoming a principle of action suggests that there is something about the motivational capacity of a passion that is relevant to determining whether or not it is calm. While there are regular connections between the motivational strength of a passion and its calmness or violence, Hume claims that the motivational strength of a passion is, strictly speaking, distinct from its being either calm or violent: "it is evident passions influence not the will in proportion to their violence, or the disorder they occasion in the temper…We must, therefore, distinguish betwixt a calm and a weak passion; betwixt a violent and a strong one" (ibid.).

There are a number of issues that arise from the strong/weak distinction. The first is that, while he is explicit that the calm/violent and strong/weak distinctions are different, Hume nevertheless says that we will ordinarily be more successful in influencing another person's actions when we cause or increase the *violence* of his or her passions.[36] A related issue is that

---

[36] "But notwithstanding this, 'tis certain that when we wou'd govern a man, and push him to any action, 'twill commonly be better policy to work upon the violent than the calm passions, and rather take him by his inclination, than what is vulgarly call'd his *reason*" (T 2.3.4.1, SBN 419). By "what is vulgarly call'd his *reason*" Hume is referring back to T 2.3.3.8 where he argues that the common view (also the rationalist view) that reason is capable of resisting violent desires with an eye towards, say, one's long

Hume does not seem to offer any explicit account of how it is that the strength of one's passions is increased, but he does offer some discussion of how passions become, or are made more, violent.[37] Also, Hume says that one of the causes of violence among passions is conflict between passions. Given this, one wonders how Hume can account for the predominance of calm over violent passions, a virtue he calls strength of mind, when the clash of passions increases their violence.[38] Finally, we might wonder whether the strong/weak distinction is meant to cut across all the other divisions of the passions in such a way that any passion could be considered motivationally strong or weak. This amounts to the question of whether or not Hume maintains that every passion is motivational or if he maintains that only certain passions, like desire and aversion, are motivational.

I will not attempt a sustained response to any of these worries. I think it is a fairly clear and important element of Hume's theory that the calm/violent distinction and the strong/weak distinction come apart, even if there are regular relations between them. I agree with Radcliffe that paying attention to the virtue of strength of mind helps to explain Hume's account here. In particular, Hume states that strength of mind, or the predominance of the *calm* over the *violent* passions, is never found in any person to such an extent that he or she will "never on any occasion…yield to the sollicitations [*sic*] of passion and desire" (T 2.3.3.10, SBN 418).[39] When a person typically possessed with strength of mind gives into such solicitations, their more violent passions are on these occasions their strong passions. However, when they are possessed of strength of mind, it is clear that their calm passions are their strong. Now, it may be that

---

term good is mistaken. Hume understands these sorts of cases as instances of the strong calm passions prevailing over weaker violent passions. It is only when we "judge things from the first view and appearance" that we mistake the calm passions for the operation of reason (ibid.).

[37] See McIntyre 2000.

[38] Baier calls attention to this problem (1991, 167-8). Radcliffe (2015) discusses several related problems with Hume's account of the virtue of strength of mind.

[39] Quoted in Radcliffe 2015, 556.

typically most people lack the virtue of strength of mind. If this were so, then it would make sense why Hume would say that it is usually more effective for us to increase the violence of the passions when we are attempting to influence other's actions. However, this response does not go far enough in explaining the casual mechanisms relating to both strength and violence, and how these mechanisms shift when strength attends the calm passions. While I think such an account is present in the *Treatise,* exploring it here would take me too far afield. The important features of the account will be discussed when I turn to Hume's discussion of moral motivation.

I must, however, say something about the final issue that I raised about the strong/weak distinction. This is the issue of whether or not Hume thinks that all passions have motivational capacities depending on their context, or whether he maintains that only certain kinds of passions, like desire and aversion, have motivational force. I think that we can gain some understanding of this issue by looking at one of the differences between pride and humility on the one hand and love and hatred on the other. Hume says that "pride and humility are pure emotions in the soul, unattended with any desire, and not immediately exciting us to action" (T 2.2.6.3, SBN 367). One might here rely on the suggestion this passage makes that it is *because* pride and humility are not attended with any desire that they fail to excite us to action. I think such a suggestion is too weak to support the claim that desire and aversion are the only motivational passions.

I do think that Hume maintains that desires and aversions are the sole motivational passions, but that his theory of the passions allows for many different passions to be typically causally linked to desire and aversion in such a way that they are ordinarily understood to play motivational roles. Love and hatred are passions that illustrate this connection. Hume says that these passions "are not compleated [*sic*] within themselves, nor rest in that emotion, which they produce, but carry the mind to something farther. Love is always follow'd by a desire of the happiness of the person belov'd, and an aversion to his misery: As hatred produces a desire of the misery and an aversion to the happiness of the person hated" (T 2.2.6.3, SBN 367). Hume is not claiming that there is a necessary, conceptual relation between love and benevolence: "I see

no contradiction in supposing a desire of producing misery annex'd to love, and of happiness to hatred" (T 2.2.6.6, SBN 368). The relation between love and benevolence is accounted for by "an original constitution of the mind" and arises in the same way that various "appetites and inclinations" arise from the "situation" and composition of the body (ibid.). Just as we might give rise in someone to a desire for food by showing him images of foods we believe him to be fond of, so too can we give rise in him to a desire for the happiness of another by turning his attention to someone we believe him to love. The motivational capacities of love, though mediated by desire, are so regularly connected with that desire that we might speak of love as being a motivating passion.[40]

Moreover, while Hume claims that pride and humility are "pure emotions in the soul," he does allow that there are certain contexts where, due to the operation of the principles of association, pride strengthens desires:

> Thus a suit of fine cloaths [*sic*] produces pleasure from their beauty; and this pleasure produces the direct passions, or the impressions of volition and desire. Again, when these cloaths are considered as belonging to ourself, the double relation conveys to us the sentiment of pride, which is an indirect passion; *the pleasure which attends that passion, returns back to the direct affections, and gives new force to our desire or volition, joy or hope*. (T 2.3.9.4, SBN 439, emphasis added)

So, even though Hume denies that pride is the proximate cause of our desire for the fine suit of clothes, he does recognize that when we have that desire, we will experience pride and this pleasurable passion will in turn give more strength to the original desire. Again, then, in certain contexts, we find that there are regular connections between passions and volitions so that we might well think of the passions themselves as having motivational force. I am here spelling out Radcliffe's suggestion that "it's true that Hume thinks actions are caused by passion, for every passion he cites as a motivating one, he also gives a definition of it in terms of desire."[41]

---

[40] McIntyre notes that this gives us reason to think that the causal origin of the direct passions is not as simple as Hume initially characterizes it (2000, 80). Here we see the desire (direct passion) arising out of the complex causal relations that give rise to the indirect passion of love.

[41] Radcliffe 2015, 548. See also Bricke 1996, 36-7.

Reason and Passions

I mentioned above that one of the important roles that the calm passions play in Hume's theory is that they are capable of being confounded with reason. This is the main element in Hume's response to what he takes to be philosophical orthodoxy. This orthodoxy is the view that there is a conflict for the government of a person between her passions and her reason, and it is reason which ought to win out for the direction of action in the prudent and virtuous. Hume of course denies this and with an all too brief account of his theory of the passions, I turn to examine his arguments for this denial.[42]

The first argument is based on Hume's understanding of reason. In general, reason is the discovery of truth and falsity (T 3.1.1.9, SBN 458).[43] Hume understands truth and falsity in terms of an agreement with what he calls "the *real* relations of ideas, or to *real* existence and matter of fact" (ibid.). Hume does not say what it is that is supposed to agree with these "real" relations or matters of fact, but I think that he has in mind an agreement between a judgment and the real relations of ideas or matters of fact.[44] Hume also does not say what it means for there to be an *agreement* between these. I think Hume means for judgments to be representational: they are copies of experiential contents. When these contents correspond to the actual conceptual relations or to real objects or events then there is an agreement. Hume argues that if reason operates independently of any passion or affection, it will never on its own be capable of producing an impulse to act.

---

[42] There are many important and difficult interpretative issues that arise from this section of the *Treatise.* In what follows, then, I think it is best to present what is taken to be a standard reading of this section. I follow Cohon 2008 in her presentation of this element of the standard reading of Hume.

[43] One of the important issues that arises straightaway is whether Hume is best understood as offering *arguments* for the claim that reason alone is motivationally inert. Cohon 1988 and 2008 points out that interpreters run into problems regardless if they read these as a priori demonstrative arguments, or probabilistic-causal arguments.

[44] I do not here intend for 'judgment' to be understood technically. I follow Hume's footnote from T 1.3.7 where he argues that "judgment and reasoning…are nothing but particular ways of conceiving our objects." Hume distinguishes these from beliefs which he maintains are "strong and steady conception[s] of any idea, and such as approaches in some measure to an immediate impression" (ibid.). Both judgments and beliefs can be true or false.

Reason can discover through its demonstrative functions that certain relations hold between our concepts. I can, for example, do some arithmetic to discover such relations. But merely arriving at a true arithmetical claim does not itself provide me any direction to act. Hume recognizes that the discovery of such demonstrative truths may, in certain contexts, take on a motivational dimension. To use his example, "[a] merchant is desirous of knowing the sum total of his accounts with any person: Why? But that he may learn what sum will have the same *effects* in paying his debt, and going to market, as all the particular articles taken together" (T 2.3.3.2, SBN 414). Discovering how much he is in debt, via calculation, does influence his actions in this context. This is presumably because he has particular ends and desires, like becoming debt free, avoiding the ire of his creditors, and etc. But if he entirely lacked these ends and desires it doesn't seem like the deliverances of his demonstrative reasoning would themselves give rise to particular actions.[45]

The case is not much better with respect to our probabilistic reasoning. The clearest case of how such reasoning could relate to our motives and actions is in our reasoning about how to, or how to best, achieve our desired ends. This will come about through experience and result in probable judgments about the causal relations between actions and their consequences. This is a major element of our mental lives since it arises whenever we think something will give us pleasure or pain. Hume says that it is "from the prospect of pain or pleasure that the aversion or propensity arises towards any object: And these emotions extend themselves to the causes and effect of that object, as they are pointed out to us by reason and experience" (T 2.3.3.3, SBN 414). The causal reasoning here depends on the prospect of pleasure and pain. If such prospects

---

[45] Baier addresses a concern with Hume's argument as it relates to comments he makes in T 2.3.10 titled "Of curiosity, or the love of truth." She argues that Hume does grant a sense in which reason alone has motivating force, and this force stems from the pleasure that we take in exercise of our rational faculties. See Baier 1991, 172. Given her reference to the pleasure and enjoyment of the *exercise* of reason, it is not clear that this point is inconsistent with the standard interpretation of Hume at T 2.3.3. There, Hume's arguments seem to focus on the motivational influence of the *deliverances* of reason, on the conclusions of deductive or probabilistic arguments. We might be motivated to engage in arguments the conclusions of which give rise to no motivational force.

are removed, then the motivational force of such inferences is entirely lost. The mere fact that there is a causal relation between two objects does not provide me with a motive to act. In order for me to be motivated by these causal judgments, I cannot be entirely indifferent to all of the objects that I understand to be causally related to them. In this way, these causal judgments *alone* cannot provide me with any impulse to act.

For Hume, then, neither of the operations of reason is up to the task of providing motivational force on its own.[46] But Hume has not yet argued that reason is *entirely* inert. One might think that, even though reason is incapable of producing a motive or volition on its own, that it might still *prevent* or *hinder* certain motives or volitions on its own. If reason is capable of extinguishing a passion, then there may be, contrary to Hume's thesis, a genuine sort of conflict between reason and (motivating) passions. Hume, however, excludes this possibility. He argues that if some principle is capable of preventing a passion, then that principle must itself be capable of giving rise to a motivational impulse. He claims that "[n]othing can oppose or retard the impulse of passion, but a contrary impulse" (T 2.3.3.4, SBN 415). Hume requires that the contrary impulse be such that, in certain contexts, it could give rise to action on its own. So, any principle that is capable of opposing a passion must be able to generate an impulse that on its own could yield action, other things being equal. Reason is incapable of this. So, Hume concludes that reason cannot generate or prevent motivational impulses to action. Because reason is inert in this way, "[w]e speak not strictly and philosophically when we talk of the combat of passion and of reason" (T 2.3.3.4, SBN 415).

A second argument has to do with the representational content of the passions. Hume says that a passion is an "original existence" and specifies that he means by this that a passion

---

[46] Since I am here attempting to present a standard reading, I intend for this claim to be ambiguous between (a) reason considered as a faculty, or a process and (b) reason considered as the deliverances of such a faculty or process. On the standard interpretation presented here, Hume means for reason in both senses to be motivationally inert. Following Baier, Cohon, and others, I should note here that if the deliverances of a process of reasoning is a *belief*, then Hume does think reason in this sense has motivational efficacy. See T 1.3.10, "Of the influence of belief," as well as Baier 1991, 171-3; Cohon 2008, 16-9.

"contains not any representative quality, which renders it a copy of any other existence or modification" (T 2.3.3.5, SBN 415). It is worth quoting the relevant passage at length:

> When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible, therefore, that this passion can be oppos'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent. (T 2.3.3.5, SBN 415)

Again, Hume maintains that having a representative capacity is necessary for anything to count as being either true or false, since truth and falsity simply consist in an agreement between a representative state, like a judgment or a belief, and whatever it is that the state represents. It is important not to confuse Hume's saying that a passion is an *original existence* with other uses of the term "original" as they relate to impressions. Recall that Hume says that impressions of sensation "arise in the soul originally, from unknown causes" (T 1.1.3.1, SBN 7). He also says that some direct passions, like benevolence, revenge, and the general desire for good and aversion to evil, arise in the same way, namely *originally*. In these latter cases, Hume seems to use the term 'original' to pick out natural operations of the body and the mind that are not susceptible to his philosophical methodology but may be susceptible to an investigation by natural philosophers and anatomists. By calling a passion an original existence, he is not trying to signal the mysteriousness of their causal origins. He seems instead to be focusing on the fact that the passions, as impressions, are not copies of other perceptions, even if their causal history presupposes previous impressions and ideas. As Cohon says, "passions, volitions, and actions as such are not signs or symbols of a reality beyond them. They do not stand for anything else. *In particular*, they are not copies of other items, as ideas are copies of the impressions that cause them (and that they consequently represent)."[47] On this reading, all passions are original existences insofar as no passions are copies of other perceptions. Some passions, further, arise

[47] Cohon 2008, 19; emphasis added.

originally from unknown causes, but not all passions do. So, some original existences do not arise originally.[48]

I think it will also help to think of Hume as using the term 'reference' in a quasi-technical sense specified within the passage. In order for a perception to refer in this sense, it must be a copy and so must admit of the sort of agreement or disagreement specified above. If we understand Hume in this way, then we avoid one interpretative challenge of this text. Hume clearly holds that the passions have intentional objects. He thinks that one of the primary distinctions between pride and humility, and love and hatred is a distinction between their objects: pride and humility always have oneself as their object, whereas love and hatred always have other thinking beings as their objects. If we understand reference in this passage to mean a lack of an intentional object, then we would have to hold that Hume either entirely forgets or revises without mention, one of the fundamental elements of his theory of the passions.[49] I think that neither option is plausible. Hume does not need to deny that our passions are directed at certain objects in order for this argument to succeed in establishing his point. All he needs to deny is that the relationship between a passion and its object is the agreement/disagreement relation necessary for truth and falsity. When I am angry with someone for an injury I believe her to have caused, my emotional state does take her as its intentional object. It is true *that* I am angry with her, and there are judgments that I make concerning her character or intentions that will be either true or false. But my anger itself is not a representation of her or of these judgments. The anger is a new impression that arises from a complex causal sequence. My judgments and beliefs do play a role in this causal sequence: my belief that the injury resulted

---

[48] Reading Hume in this way helps to avoid a problem with the argument from the representational content of the passions. Baier argues that Hume's argument here is "very silly." I think that she makes this claim, in part, because she identifies a perception's arising originally with its being an original existence. If we accept Baier's identification, then Hume is straightforwardly inconsistent. He calls passions both original existences (arising without any previous perceptions on Baier's reading) and secondary impressions (which depend on previous perceptions). See Baier 1991, 160.

[49] This is another reason why Baier thinks the argument from representation is silly. For the reasons I specify here, I do not agree with this interpretation.

neither from her character nor her intention will quell my anger. In this case, we would not say that my anger was false or that it misrepresented its object. My anger was directed at someone who did not intend to cause me injury and does not have a blamable character, but the perception itself was not a representation or a copy of her. So, the impression itself could be neither true nor false.[50]

This illustrates that reason does play some indirect role in our passions and our actions. Recall, reason is and ought only to be subservient to the passions. This subservient role allows us to speak of certain passions or actions as being *unreasonable* in a loose and unphilosophical way. There are two ways in which passions or actions can be, in this loose sense, unreasonable, and both of them involve false judgments or beliefs. The first way involves a false judgment or belief about the existence of an object. I may judge that a beeping noise is coming from a ticking time bomb in the next room over, and so become fearful. Insofar as there is no bomb my judgment is false, and my fear unreasonable. But the fear is not *itself* unreasonable since it is a nonrepresentational state. The fear is only in a sense unreasonable because it is accompanied by a false judgment that, because it consists in a disagreement with a matter of fact, is strictly unreasonable. The second loose sense in which a passion or action can be unreasonable involves false judgments about the relations of cause and effect. Suppose that I falsely believe that it is far less likely for a smoker to develop lung cancer if he continues to smoke than it is for a former smoker to develop lung cancer once he quits. This belief strengthens my desire to smoke (or my aversion to quit) and this in turn influences my actions. It is unreasonable for me to continue smoking but only because I have a made a false judgment about the better means to avoid lung cancer. My desire to continue to smoke is not itself unreasonable. Without a false judgment, my desires are neither reasonable nor unreasonable.

This may sound rather counterintuitive. Isn't smoking cigarettes one of the most unreasonable things a person can do? Suppose that, as a smoker, I believe that the short term

---

[50] For more on the kind of reading that I am sketching here, see Weller 2002 and Bricke 1996.

pleasures of smoking are greatly outweighed by the long term pains of succumbing to one of the many illnesses (often terminal) to which smoking likely gives rise. Suppose, further, that my desires are still stronger for the short term, immediate pleasures. Isn't this a paradigmatic case of an unreasonable desire, and aren't we inclined to say that acting in light of these desires is also unreasonable? Since this case does not depend on my making a false judgment, Hume is committed to the view that neither the passion nor the actions to which it gives rise are unreasonable. He says that it is not "contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter" (T 2.3.3.6, SBN 416).[51] Hume acknowledges that this view is "somewhat extraordinary" (T 2.3.3.4, SBN 415). However, it is meant to be a straightforward consequence both of his account of the operations of reason and his theory of the passions. The thesis is counterintuitive and Hume is aware of this. He would likely point out that believing that preferring our own acknowledged lesser goods is irrational is a consequence of the aforementioned philosophical orthodoxy:

> Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and assert that men are only so far virtuous as they conform themselves to its dictates. Every rational creature, 'tis said, is oblig'd to regulate his actions by reason; and if any other motive or principle challenge the direction of his conduct, he ought to oppose it, till it be entirely subdu'd, or at least brought to a conformity with that superior principle. (T 2.3.3.1, SBN 413)

So, to the charge that his view is counterintuitive, Hume would likely respond by pointing out that it seems so only from the perspective of this pervasive, yet mistaken, philosophical tradition.

One might think that the orthodox view of the combat between reason and passion has an experiential advantage. There are many cases, though perhaps some of us might wish there were more, where we do not behave as the smoker does in the above discussion and do not develop stronger desires for our acknowledged lesser, short term goods. In these cases, it seems that we are using our reason, which tells us the direction of our greater good, and we are resisting our

---

[51] I will have more to say about Hume's claim here in the following chapter.

desires that are out of proportion with what our reason tells us. That is, these seem to be cases where we experience a combat of reason and passions, and, what is more, cases where reason is the superior and guiding principle. Hume, however, responds that this is an inaccurate description of the experience, heavily influenced by the mistaken philosophical orthodoxy. He argues that these experiences depend on conflating two mental phenomena because of their similar phenomenology: Again, since the calm passions and reason both operate without "any sensible emotion," we mistake the operations of the calm passions with those of reason "because their sensations are not evidently different" (T 2.3.3.8, SBN 419).

Necessity and Will

I have shown above that there are some important threads connecting some of Hume's discussions through all three books of the *Treatise*. One of the most obvious threads is that linking Hume's denial that morality is grasped through reason alone given the role that morals have on our passions and actions and given that reason alone could never play such a role. I turn to briefly examine Hume's account of how the passions motivate.

It is well known that Hume argues against a conception of human will whereby it is free in the sense of lacking determining causes. He rejects that the will has liberty of indifference where this is understood as "a negation of necessity and causes" (T 2.3.2.1, SBN 407). One of the central elements of his argument for this claim is that there is no difference between the causal relations of necessity that are understood to hold between physical objects and the causal relations of necessity that are thought to hold between motives and actions. Hume offers two definitions of necessity that he takes to correspond to the two definitions of cause that he offered in Book 1: "I place [necessity] in the constant union and conjunction of like objects, or in the inference of the mind from the one to the other" (T 2.3.2.4, SBN 409). Hume argues that moral evidence, his way of referring to an inference from motives to actions, arises from regular unions with an irresistible movement from one idea to the other. For example, he claims that such causal

reasoning "mixes itself so entirely in human life, that 'tis impossible to act or subsist a moment without having recourse to it" (T 2.3.1.15, SBN 405). He continues:

> A prince, who imposes a tax upon his subjects, expects their compliance. A general, who conducts an army, makes account of a certain degree of courage. A merchant looks for fidelity and skill in his factor or super-cargo. A man, who gives orders for his dinner, doubts not of the obedience of his servants. In short, as nothing more nearly concerns us than our own actions and those of others, the greatest part of our reasonings is employed in judgments concerning them. Now I assert, that whoever reasons after this manner, does *ipso facto* believe the actions of the will to arise from necessity, and that he knows not what he means, when he denies it. (Ibid.)

It might be wondered how exactly these cases are meant to illustrate *necessity*. Hume is well aware that sometimes servants are disobedient, sometimes soldiers flee a battle, and sometimes citizens do not pay their taxes. These cases are not intended to show that humans will always do as those with authority over them command, since this is not what Hume means when he talks about necessity.

Hume is trying to establish that there are certain regular connections between interpersonal actions, like a general's commands and his soldier's obedience with those commands. In our experience the one is regularly conjoined with the other, and the thought of the one is naturally carried to the thought of the other. If as a captive I overhear the general of my enemy command his soldiers to kill me at dawn, I do not wait until I am on the gallows to see if they will follow his orders, but immediately make my final attempts to escape. Now, Hume allows that my judgments about these relations come in degrees. The more frequent my experience of the conjunction, the more probable my belief. So, the mere fact that some soldiers may refuse their general's orders to kill their captives does not show that there is no regular connection.[52] What is more, when there is a lack of a regular connection on a particular occasion, this does not support Hume's opponent. We will not say that there was no cause or

---

[52] Hume says, "When any phenomena are constantly and invariably conjoined together, they acquire such a connection in the imagination, that it passes from one to the other, without any doubt or hesitation. But below this there are many inferior degrees of evidence and probability, nor does one single contrariety of experiment entirely destroy all our reasoning" (T 2.3.1.12, SBN 403). See also Norton 2000, I67-8.

motive for why a soldier deserted a battle. And, in fact, one of the things that an astute general is aware of is the character of his soldiers, and this is important for him to pay attention to so that he will be better suited to judge who is likely to run and who is likely to fight.

This last point about knowing the character of those with whom one is involved brings out an important element in Hume's discussion of the will. I mentioned above that Hume was eager to show regular connections between interpersonal actions, at least that is what it seems he is doing with the examples provided at T 2.3.1.15. I, of course, do not here mean to make Hume out to be proto-behaviorist, since he is not. He is not merely trying to show that behavioral regularities exist when certain kinds of beings are placed in certain experimental conditions. The nature of behavior, considered in itself, makes it the wrong object of study for Hume's new science of man. When Hume responds to those who maintain that we love or hate individuals in proportion to their intentions to give us pleasure or pain, he says that knowing the intention of another is not necessary for these passions, so long as we can see her pleasure- or pain-causing actions as resulting from her "person and character" (T 2.2.3.4, SBN 348). It is only when we cannot connect a benefit or an injury to another's character that we need to connect the benefit or injury to his intention if we are to feel love or hatred towards him.[53] "'Tis not enough, that the action arise from the person, and have him for its immediate cause and author. This relation alone is too feeble and inconstant to be a foundation for these passions. It reaches not the sensible and thinking part, and neither proceeds from anything *durable* in him, nor leaves any thing behind it; but passes in a moment, and is as if it had never been" (ibid.).

Hume returns to consider the same position in connection to his discussion of the causation and necessity of the will. Another major element of his argument is to claim that attributions of responsibility and institutions of punishment do not square with notions of justice

---

[53] Note, however, that Hume does recognize that hatred and anger, or a desire for harm to another, may arise from any injury regardless of the character or intention of the injurer. We may first feel the passion, then seek after the causes. This is no counterexample to what I am claiming on behalf of Hume. The idea is that, so long as the injury stems from the action alone, without reference to character or intention, the passion will be extinguished. See T 2.2.3.6-10.

and morality if the will has liberty of indifference. As Hume understands the indifference view, it entails that no actions are caused by general character traits or by an individual's intentions, since no actions are caused at all. This means that attributions of responsibility attach merely to actions considered in themselves. Hume is clear that actions are too fleeting and feeble to be the basis of our passions, and he extends this view to claim in addition that actions are too fleeting and feeble to be the objects of responsibility and punishment. As further confirmation of his thought, he points out that we do not blame people for evil acts done in ignorance.[54] He claims that this is because "the causes of these actions are only momentary, and terminate in them alone" (T 2.3.2.7, SBN 412). This helps clarify Hume's point. Hume thinks that punishment based on accurate judgments of responsibility are important insofar as they help shape future behavior.[55] Future behavior can only be affected through punishment if it operates on features of individuals that are durable, not fleeting. We need to see individuals' actions as arising from relatively stable sets of characteristics so that we can influence them towards characteristics more regularly conjoined with desirable behaviors. So, Hume's principal concern is not to establish a regularity of interpersonal behaviors, but is instead to establish a regularity between characters and motives, on the one hand, and behaviors on the other.

One final comment on Hume's account of freedom and necessity is in order. I have not yet said what exactly Hume takes the will to consist in. Hume says that he understands the will as "*the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*" (T 2.3.1.2, SBN 399). Hume claims that the will is simple and indefinable, like the passions of pride, humility, love, and hate. This suggests that the will is, like these passions, a secondary impression. It would fit within Hume's theory to say that the will depends on other perceptions, such as the impressions or ideas of pleasure and pain. For example, I feel the impression that constitutes the will when after I have felt pleasure,

---

[54] Hume is not explicit about this point, but it is clear that he means ignorance that is itself non-culpable.

[55] Baier 1991, 154-5.

either directly from some object or indirectly through the belief that the object will give me pleasure, I have further impressions of sensations from the kinesthetic sense of my body moving towards the pleasurable object.

It is clear that Hume understands that the will is a sort of secondary impression. However, this does not stop him from talking sometimes as if the will were a faculty. He says, for instance:

> The will exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body. Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T 2.3.9.8, SBN 439)

I am not here attempting to provide a thorough interpretation of the will. However, I do think that when Hume talks of the will *exerting* itself, what he means is that we feel an additional secondary impression whenever pleasure pulls us towards or pain pushes us away from some object. This is consistent with Hume's claim that the "chief spring or actuating principle of the human mind is pleasure or pain" (T 3.3.1.2, SBN 574). Of course, I do not need to act from a view of obtaining pleasure, or avoiding pain, since, as we see in this passage, Hume recognizes that certain sorts of instincts arise from "unknown causes" such as the desire for happiness for one's beloved, revenge, lust, thirst, and etc. In these cases too the will is felt as a result of certain passions being present, and so would still be considered a secondary impression.

Moral Evaluation

Hume's theory of the will and its relation to motivational passions is connected to his account of moral evaluation.[56] For Hume, actions are not the primary locus of moral evaluation. This is because, as we have already seen, Hume thinks that actions are fleeting and while their

---

[56] Tate says, "Hume commits himself to the principle that we evaluate acts on the basis of their motives prior to his discussion of justice. He lays the groundwork for this claim in Book II in his discussion of love and hatred, and *liberty and necessity*" (2005, 100).

consequences may be either pleasurable or painful, we do not have stable emotional responses to fleeting causes. But Hume does think that our moral evaluations are durable and lasting and so reflect durable and lasting elements of the objects of our evaluations: he says that our moral evaluations "depend upon durable principles of the mind, which extend over the whole conduct, and enter into the personal character" (T 3.3.1.4, SBN 575). So, Hume says that when we evaluate actions "we regard only the motives that produc'd them, and consider the actions as signs or indications of certain principles in the mind and temper" (T 3.2.1.1, SBN 477). Hume's claim that character traits and motives are the primary locus of moral evaluation thus stems directly from his understanding of the relationship between motives and action. This additionally provides Hume with a relatively stable basis for moral claims. Hume insists that there are regular connections between certain motives and certain behaviors, and that observers must make inferences about behavior on the basis of motives if they are to operate in society with other people. The more uniformly we've experienced a behavior following from a character, the stronger our judgment of their causal relation. So, the regular connections that Hume points out in Book 2.3 as pervading "politics, war, commerce, œconomy" turns out to be necessary for moral evaluations as well (T 2.3.1.15, SBN 405).

It is important not to be too hasty in understanding the conclusions of Hume's argument in Book 3.1. We have not yet examined what Hume means when he says that moral distinctions naturally have an influence on the passions and actions. We have seen why Hume's theory of the passions and of reason leads him to deny that reason alone has any such influence. These two claims together might be understood to entail that Hume denies that moral evaluations are capable of being either true or false. It must be admitted that Hume's language at certain points helps generate this reading. When Hume returns to consider the argument about the impotence of reason alone in Book 3 he says:

> Reason is the discovery of truth or falshood [*sic*]. Truth or falshood consists in an agreement or disagreement either to the real relations of ideas, or to real existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now it is evident our

> passions, volitions, and actions, are not susceptible of any such agreement or disagreement; being original facts and realities, compleat [*sic*] in themselves, and implying no reference to other passions, volitions, and actions. It is impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason. (T 3.1.1.9, SBN 458)

Hume later paraphrases the conclusion of his 3.1 argument by saying that morality "is more properly felt than judged of" and he goes on to describe our moral feelings in a way that makes them very much like the calm passions: "this feeling or sentiment is commonly so soft and gentle, that we are apt to confound it with an idea, according to our common custom of taking all things for the same, which have any near resemblance to each other" (T 3.1.2.1, SBN 470). If moral evaluations are passions, and passions do not admit of truth and falsity, Hume seems committed to the view that our moral evaluations also do not admit of truth and falsity. There is then, it seems, a very natural way of reading Hume as a metaethical noncognitivist.[57]

Appealing to these texts alone as providing support for a noncognitivist reading of Hume poses interpretive difficulties. The most glaring problem is that Hume seems to claim that moral judgments have cognitive content. In a rather famous passage, Hume says:

> Take any action allow'd to be vicious: Wilful [*sic*] murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions, and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You can never find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. *Here is a matter of fact*, but 'tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, *you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it*. (T 3.1.1.26, SBN 468-9; emphasis added)

This passage suggests that Hume thinks moral evaluations are matters of fact regarding the feelings that arise from the constitution of human nature. As Richard Fumerton puts it, "Hume has apparently found a matter of fact with which to identify the subject matter of moral

---

[57] For problems that arise when we treat this argument as an argument for noncognitivism, rather than an argument for the claim that actions and passions are not the right kind of ontological items to count as being reasonable or unreasonable, see Cohon 2008, 19-23.

judgment."[58] I will return below to briefly address the ambiguity that arises in this passage between understanding the relevant judgments of matter of fact as descriptions of what all or most people feel upon the contemplation of "wilful murder," and understanding the relevant judgments of matter of fact as descriptions of what an individual feels upon such contemplation.[59] This issue notwithstanding, Fumerton notes that Hume here seems to be asserting an "unequivocal statement describing the meaning of ethical terms," but he also recognizes that there are some passages where "Hume seems to be moving toward a version of noncognitivism."[60] If we simply look at the passages that favor noncognitivism and compare them to those passages that favor a certain kind of descriptivism, the question of which view Hume would have adopted begins to look intractable. It may in fact be the case, as Fumerton suggests, that Hume was simply confused about the different metaethical positions his statements seem to describe. Ultimately, Fumerton argues that, even though Hume was trying to develop either a noncognitivist or a certain kind of subjective descriptivist account, "everything fell into place for Hume" in the willful murder passage and this, in part, leads him to favor a certain kind of subjective descriptivist account.[61] I think many interpreters attempt to settle debates about Hume's metaethics by championing those passages where he or she thinks "everything falls into place for Hume," and Fumerton's candor here is to be admired. Other interpretations, however, attempt to locate considerations in favor of noncognitivism in claims regarding the consistency of Hume's overall system of moral philosophy. I turn to one such argument presently.

---

[58] Fumerton 1990, 35.

[59] Ibid., 35-6, and 75-82.

[60] Ibid., 75-6.

[61] Ibid. 76-7. Since I am not attempting to settle any of these metaethical disputes here, it will be enough to present Fumerton's "crude" accounts of subjective descriptivism: (I) "To say that something is good is to say that most people value (approve of) it"; (II) "'X is good' said by S means the same as 'I value (approve of) X' said by S." (1990, 38).

An important argument for adopting a noncognitivist interpretation of Hume arises from what Smith calls "The Moral Problem," and what Sayre-McCord and Radcliffe call "The Metaethical Problem."[62] The Metaethical Problem arises from a triad of seemingly inconsistent claims. The first is a commitment to what has come to be known as a Humean theory of motivation. On this theory, belief and desire are different psychological states and both are required for motivation. The second claim is a commitment to an appraiser version of motive internalism: it is necessarily the case that if someone judges that she ought to do something, then she has some motive to do as she judges. The third claim is a commitment to moral cognitivism: moral judgments are true or false because they "impart information of some sort."[63] If moral judgments are cognitive then they fall on the side of beliefs in the Humean belief/desire distinction. The Humean theory of motivation holds that beliefs alone do not motivate, but that desires must also be present. There is, then, no guarantee that if an agent morally judges that she should act in a particular way that she will thereby be moved to do so.[64] The three claims do not seem to be able to be all true. The idea here is that if Hume accepts the Humean theory of motivation and if he accepts internalism, then he cannot accept cognitivism. If he were to hold that moral judgments are cognitive, then he would either have to give up his claim that cognitive states, like beliefs, do not alone provide motivation, or he would have to give up his commitment to internalism. Since many have seen these latter claims are fairly clear commitments of Hume's theory, and because of the textual evidence that suggests noncognitivism, the thought is that Hume should not be taken as a cognitivist.

---

[62] Smith 1994, 1-15; Radcliffe 2006, 353-5.

[63] Radcliffe 2006, 353.

[64] As Radcliffe develops this problem, it is not clear how it presents a difficulty for subjective descriptivist interpretations. I will say a bit more about these readings below. The view is, roughly, that moral judgments are descriptions of appraiser's attitudes. Depending on how one specifies these attitudes, one seems to be able to adopt all three claims without conflict.

Radcliffe argues, however, that Hume does in fact accept all three claims. I will take for granted that Hume's theory of motivation is sufficiently similar to the Humean theory that whatever differences there may be will not present a problem for understanding his response to The Metaethical Problem. I argue in the next chapter that there are some reasons to favor an internalist reading of Hume. Given these commitments, how could Hume accept a version of cognitivism? The key to Radcliffe's interpretation is Hume's account of belief and his account of judgment.[65] Hume maintains that "'Tis certain, that the belief superadds nothing to the idea, but only changes our manner of conceiving it, and renders it more strong and lively" (T 1.3.8.7, SBN 101). A belief is just an idea that is made livelier by the presence of impressions. When I believe that my dog is by the door, I have an idea of my dog being by the door and this idea is made livelier by the impressions that I am currently having. The idea and the belief have the same content, what is different about them is that the latter is more lively and vivid due to other present perceptions.

Radcliffe adopts the standard reading of Hume's theory of perceptions that holds that ideas are representative but impressions are not.[66] What this allows Hume to do is to draw a distinction between beliefs and the causal processes by which they are formed. He can distinguish *the process of judging* from the *product of judging* (i.e. the judgment).[67] Hume's account of beliefs makes it clear that he thinks beliefs have two distinguishable aspects. There is the aspect that a belief has in common with an idea. The idea, as a representation or copy, is cognitive because it is the sort of thing that admits of agreement/disagreement with relations of ideas and matters of fact. There is therefore a cognitive aspect of belief as well. Hume also thinks

---

[65] In what follows, I present Radcliffe's interpretation of Hume's account of belief and judgment. See Radcliffe 2006, 359-61.

[66] Ibid., 359. Don Garrett offers a different interpretation whereby impressions are representational. See Garrett 2006. As I argued above, the claim that impressions are not representational does not preclude Hume from saying that the passions have objects.

[67] For a similar distinction see Cohon 2008, 14.

that there is a noncognitive, nonrepresentational aspect to belief that it does not share with ideas. Beliefs have a "force and vivacity" that mere ideas lack. The process of coming to believe, for Hume, crucially depends on the operation of noncognitive processes. As an example, Radcliffe points to our coming to believe that a causal connection obtains between two events. When I believe that A causes B, I have a feeling of anticipation or expectation that B will occur when I have an idea of A occurring, and this expectation arises from my having experienced the constant conjunction of A's and B's. The enlivening impression that turns my idea of A causing B into my belief that A causes B is this feeling of expectation (rather than any impression of sensation). The process of forming a judgment, a belief, then crucially depends on there being some noncognitive process. So, Radcliffe says, "there is nothing in the process of moral judging that precludes its outcome from being an idea (and therefore, cognitive) as well, given that the other types of judgment involve a non-cognitive mental change that leads to a cognitive result."[68] As I argue in the next chapter, Hume takes the noncognitive process that produces moral judgments to depend on motivating sentiments.

I do not think that this consideration settles the issue regarding cognitivism and noncognitivism in Hume. Even if it did, however, there are still a number of other related metaethical questions that would need to be settled.[69] For instance, does Hume think that our cognitive moral judgments involve the projection of psychological fictions onto characters and actions so that he ends up defending an error-theory?[70] Or perhaps he thinks that such judgments are descriptions of the sentiments of spectators? Following Fumerton, we should note that this last suggestion might be developed in a couple of different ways. First, the moral evaluation is a

---

[68] Radcliffe 2006, 361.

[69] In what follows, I follow Cohon's presentation of the interpretive options that are on offer. See Cohon 2010.

[70] Mackie thinks there are two plausible interpretations of Hume, and this "Objectification view" is one of them. See Mackie 1980, 73-4. This view is cognitivist insofar as it holds that moral judgments are truth-evaluable, but, since all moral judgments depend on projecting psychological fictions, they all turn out to be false.

description of the actual sentiments of an observer so that the claim "subtle pride is a virtue" *when said by S* means nothing other than "I have a feeling of approbation on the contemplation of subtle pride" *when said by S.*[71] Alternatively, the claim that "subtle pride is a virtue" just means that most people approve of subtle pride.[72] A related interpretation holds that when we evaluate motives and traits as being morally praiseworthy, we ascribe to that action a property such that it would give rise in impartial and sympathetic spectators to a distinct feeling of moral approbation. On this sort of a view, moral judgments are cognitive because they report on facts about the dispositional tendencies of motives and character traits.[73] In addition, some interpretations distinguish two different psychological states that Hume can call moral evaluation. This seems to be the interpretation offered by Radcliffe above. On the one hand, there is there a feeling of approbation or disapprobation and on the other there is a belief or judgment. The former is noncognitive and the latter cognitive. The question then arises what the content of the cognitive state is supposed to be, and, of course, commentators disagree.[74]

It is clear that settling the issue in favor of cognitivism still leaves unanswered a number of questions about Hume's theory of moral judgment. One of the overall conclusions of my project is that interpreting historical figures in light of contemporary taxonomies is a very difficult task and there is no guarantee that the actual views under consideration will fit neatly within contemporary frameworks. This is not to say that such attempts are unhelpful or doomed to failure. It does suggest, however, that such attempts must be undergone with a great deal of care. Since my focus in this dissertation is on locating the sorts of internalist commitments in Hume (and the other empirical naturalists), I would be doing myself a disservice if I rejected out-

---

[71] Fumerton favors this reading. I have not provided his arguments for this reading here. See Fumerton 1990, 36; 77-82.

[72] Stevenson (1937) 1970, 256.

[73] This is the other interpretation that Mackie finds plausible.

[74] Capaldi 1989, 126; Cohon 2010; and 2008, chapter 4.

of-hand any attempt to place historical figures in contemporary lights. I think that such attempts are worthwhile, both for what they teach us about the historical figures and about the contemporary lay of the land. My own attempt is focused on the question of internalism. It is certainly the case that noncognitivism fits rather well with internalism, and so it might be thought that my job would be much easier with respect to the internalism question if I first established Hume's noncognitivism. But, first, there is no in principle reason why a noncognitivist could not also accept a version of externalism, and, second, establishing Hume's commitment to noncognitivism seems no less daunting a task than establishing his commitment to internalism. I will not, therefore, continue to examine where Hume falls with respect to these other contemporary metaethical theses.[75]

[75] In some of what follows, I do lean towards a cognitivist interpretation that is not a version of subjective descriptivism in either sense. I leave arguments for this presumption to a full account of Hume's moral theory, which I do not aim to provide here.

CHAPTER 6

INTERNALISM AND OBLIGATION IN HUME

Introduction

The central focus of this chapter will be on locating a discussion of ethical internalism in Hume, relating this to Hume's discussions of moral obligation, and seeing how this all relates to what Hume does or can say about the "why be moral" question. In the first section, I argue that Hume's Motivation Argument against the rationalists presupposes a version of internalism. In the second section, I introduce a passage that I call the Sense of Duty Passage that is another one of the main areas of contention in debates about internalism in Hume. In this section I also introduce Hume's conception of moral obligation. I then move on to an extended argument in favor of reading Hume as an internalist. I focus on interpreting the Sense of Duty Passage and argue that there is some reason to think this passage shows that Hume holds moral sentiments to be themselves motivational. However, as we will see, the textual support for this claim is not as robust as one would hope, and I point to a plausible externalist reading of this passage. I argue that charity requires us to favor the internalist reading.

The next main section of this chapter concerns a conceptual problem that seems to arise from Hume's maintaining a version of internalism and his holding that moral sentiments only arise from the common point of view. I offer a couple of solutions to this apparent problem.

Finally, I address the resources that Hume has available to answer the question "why be moral?" Hume's insistence that passions and actions are neither reasonable nor unreasonable means that he can only understand this question as a question about motives, and since he thinks moral judgments are motivating states, one's accepting a moral obligation entails that one has at least some motive to act according to that obligation. However, if we are looking for a motive to occupy the moral point of view, Hume is happy to say that his moral theory cannot provide us with one and that what we need is to be inspired, perhaps through great works of art, to take up the common point of view.

Internalism and the Motivation Argument

There has been much debate about the question of internalism in Hume. The main set of issues centers on reconciling Hume's commitments in one of his initial arguments against the rationalists with other elements of his moral theory that seem to conflict with these commitments. The natural place to begin an investigation into this debate, then, is to locate the commitments that stem from Hume's anti-rationalism. I have briefly mentioned the relevant argument in the previous chapter, but it will here be helpful to name the argument and necessary to examine it in some detail. Hume's *Motivation Argument*, can be usefully paraphrased as follows:[1]

1. "Morals…have an influence on the actions and affections"
2. "Reason alone…can never have any such influence"
3. "It follows, that [morals] cannot be deriv'd from reason…" (T 3.1.1.6, SBN 457).

In the previous chapter, I explained why Hume accepts premise 2: the truth or falsity of a belief does not provide motivation independent of one's concern for that which the belief is about. Premise 1 will be the focus of the discussion here. As I mentioned previously, there are at least two different ways that morals may have an influence on the actions and affections. I will here present what I, following a number of others in the literature, take to be some compelling reasons for taking one of these two interpretations seriously.[2] Before I do so, I consider what Hume means by the terms 'morals' in this premise.

Perhaps when Hume says that 'morality' and 'morals' have an influence on human passions, actions, and affections, what he means to be suggesting is that morally virtuous *traits* have such an influence. For Hume, the moral worth of an action depends on the moral worth of the character trait from which that action flowed. These traits are not intrinsically virtuous or

---

[1] This paraphrase is due to Cohon. See Cohon 2008, 13.

[2] Brown 1998, 69-87; Coleman 1992, 331-347; Radcliffe 1996, 383-407; Foot 2003; Mackie 1980. This interpretation is part of the view that Cohon calls the "common reading" of Hume. See Cohon 2008.

vicious. Their virtue or vice is determined by an observer's reaction after contemplating them under conditions of impartiality. The contemplation of pleasant and useful traits, considered from a disinterested perspective, gives rise in observers to a particular sort of pleasure and this "feeling or sentiment…denominates it [the trait] morally good" (T 3.1.2.4, SBN 472). When I consider your generous action from a disinterested perspective the mechanism of sympathy gives rise in me to a pleasure that Hume identifies with moral approbation.[3] So, when Hume says that morals have an influence on human passion and action, and he means by this that traits denominated virtuous have an influence on passions and actions, the claim is obviously true. Virtues like generosity and courage obviously influence actions, and our denominating them virtues shows, on Hume's view, their effects on an observer's passions.

I do not think this is what Hume means by 'morals' in Premise 1. If this is what Hume means, then he would have thought that the rationalists were committed to the view that the moral virtues themselves were simply conclusions of reason. If we use a particular virtue as an example it will help to make this point. The motivation argument would then say that a particular virtue, like generosity, influences human action and passion. This is certainly true, and given Hume's understanding of the role and limitation of reason alone, the second premise states that reason alone has no such influence. The conclusion would then be that generosity is not a conclusion of reason alone. This is perhaps a sound argument but it is a sound argument against a view that Radcliffe rightly calls "incredible."[4] On this interpretation, Hume must have thought that the rationalists were committed to the view that individual virtues were purely intellectual

---

[3] As I understand him, Hume does not think that every sympathetic response counts as a feeling of moral approbation or disapprobation. These sentiments must be generated from particular sorts of perspectives. This suggests that moral approbation and disapprobation are sentiments that can be identified only by their causal origins. However, there are passages where Hume seems to hold, à la Hutcheson, that approbation and disapprobation are unique *feelings* distinguishable through introspection (T 3.1.2.4, SBN 472). As we saw above, this tension also arises for Hume's account of the indirect passions. See the chapter 5 called "The Multiple Divisions of the Passions" for my sketch of a proposed solution to this concern.

[4] Radcliffe 1996, 388 n.13. I follow Radcliffe in rejecting this interpretation for this reason.

endeavors, not only in their epistemic derivation, but also in their operation. There is no evidence that Hume held this false view about the nature of the rationalist's position.

It is important to recall that Hume takes the primary issue between himself and the rationalists to be an epistemological issue: "*Whether 'tis by means of our* ideas *or* impressions *we distinguish betwixt vice and virtue, and pronounce an action blame-able or praise-worthy*" (T 3.1.1.4, SBN 457). This gives us some reason to think that by 'morals' and 'morality' he means to refer to something like a moral judgment or moral belief. If the nature of our moral judgments and beliefs is what is at issue, then it is only natural for Hume's Motivation Argument to begin with a claim about the influence such judgments or beliefs have on human passion and action. I think that this reading is right as far as it goes, but it's not yet entirely clear what a moral judgment or belief is for Hume. As I suggested in the previous chapter, this issue is quite contentious in Hume scholarship. I cannot hope to settle this question here. Fortunately, I do not think that very much of my discussion in the following section hangs on any particular interpretation of Hume's account of moral belief and judgment. Regardless of how one understands Hume's view, the Motivation Argument is making a claim about the influence of these sorts of things whatever they are.

Let's suppose that when Hume says that morality has an influence on a person's actions and affections that what he means is that if an individual judges that she ought to do something, then she typically has at least some motivating passion to act in accordance with her judgment. Suppose that she has a standing desire to do that which she judges she morally ought to do. When she judges that she morally ought to do something, her judgment informs her that the object of one of her desires is present and so the desire is capable of being satisfied. On this view, a moral judgment influences our actions and affections in the same way that our beliefs about other matters of fact influence our actions and affections: these beliefs and judgments inform us of the existence of the objects of our standing desires. On this view, my belief that there is coffee in the pot leads me to act in virtue of my desire to have a cup of coffee. If I do not desire to have a cup of coffee, then my mere judgment that there is coffee in the pot will not lead

me to get a cup of coffee. Similarly, my judgment that some course of action is morally right leads me to act in virtue of my desire to do what is morally right. If I have no desire to do what is morally right, then my mere judgment that some course of action is morally right will not lead me to act as I judge that I ought. According to this view, the moral judgments themselves do not provide motives to act. These are supplied by the independent standing desire to do what one judges she morally ought to do. This view is a version of ethical externalism because it is in principle possible for one to have a moral judgment but nevertheless fail to be moved.

Alternatively, when Hume says that morality influences our actions and our passions, he might mean to be endorsing a much stronger thesis regarding the nature of this influence. He might mean that our moral judgments are motivational in themselves. As some commentators who read Hume as an internalist say, the conditions under which an agent judges that she ought to perform some action are the same as the conditions under which an agent has a motive to perform that action.[5] Their idea here, while not clearly spelled out, is that these conditions are identical in virtue of a necessary connection between moral judgment and motivation. This is a version of appraiser motive internalism, and I will return below to consider it in more detail. For now, I want only to contrast it with the version of externalism suggested above.

On the externalist reading, Hume's argument against the rationalists is not valid. The externalist reading maintains that morality is motivationally relevant only to the extent that it informs us that one of our independent desires may be satisfied. If we grant the second premise, then reason alone cannot motivate. But this is entirely consistent with maintaining that reason discerns the distinction between virtue and vice.[6] It is entirely possible, given the externalist reading, that reason informs us of the distinction between virtue and vice on some particular

---

[5] I intend for this to be a general characterization of internalism as it is discussed by Radcliffe and Kalt. Both Radcliffe and Kalt note their indebtedness to Korsgarrd's characterization of internalism in informing their own. I do not intend to here convey all the detail and subtlety of Radcliffe's, Kalt's, or Korsgarrd's views of internalism. I intend only to offer a minimal sketch of the view in order to contrast it with the externalist reading above.

[6] For an analogous argument, see Radcliffe 2006, 356.

occasion, and given an independent moral desire, we are influenced to act in accordance with them. As Charlotte Brown helpfully points out, the externalist reading separates the epistemological question from the motivational question in such a way that the motivational influence of morality becomes irrelevant to its epistemological source.[7] Therefore, many have thought that Hume must accept a version of internalism because otherwise his argument against the rationalists is invalid.

Another helpful way of seeing the issue here is to note that the externalist reading maintains that moral judgments influence action only by connecting up with passions and desires that agents already have. As we have seen, Hume maintains that reason is capable of connecting with passions and desires in the same way. We saw that in the section titled "Of the influencing motives of the will," Hume says, "passions can be contrary to reason only so far as they are *accompany'd* with some judgment or opinion" (T 2.3.3.6, SBN 416). What is more is that our judgments can influence our desires in this way: "I may desire any fruit as of an excellent relish; but whenever you convince me of my mistake, my longing ceases" (T 2.3.3.7, SBN 416-7). Hume nowhere asserts that reason is wholly unconnected with passion or action. His claim is that *reason alone* has no such influence. Since it is the case that Hume recognizes that reason can influence passions and actions by connecting up with or accompanying existing passions, if this is the sense in which Hume thinks that morality influences passions and actions, then Hume cannot infer that moral distinctions are not discerned through reason. Reason could be the epistemic faculty through which moral judgments are made, and then these judgments and beliefs as products of reason can connect up with existing passions and desires in such a way that they influence action.

So, it looks as though Hume is committed to accepting the view that the motivating influence of moral judgments does not depend on these judgments connecting with,

---

[7] Brown 1988, 75. Brown argues that while Hume needs to defend a version of internalism to make his argument here valid, he does not develop such an account in the subsequent sections of the *Treatise*.

accompanying, or "triggering" passions and desires that agents already have. There are various ways of spelling out this requirement in the literature. Some say that Hume is committed to the view that moral notions, presumably including judgments, must *themselves* provide motivation through the moral feelings of approval or disapproval.[8] Some others say that Hume is committed to maintaining an "internal connection" between moral judgment and moral motivation.[9] Still others, trying to highlight Hume's departure from the rationalists, require that Hume specify a necessary connection between moral evaluation and motivation.[10] Whether or not these requirements amount to the same thing is unclear. What is clear, however, is that the Motivation Argument does not succeed if he understands the influence that morality has on passions and actions to be the same sort of influence that he elsewhere claims that reason has on passions and actions. The connection between morality, motivation, and action must be something different.[11]

Hume's Circle and the Sense of Duty

Because Hume's discussion in T 3.1.2 does not relate to the motivational role of the moral sentiments, he has not yet developed an internalist account when he moves out of part 1 of Book 3 and into part 2, "Of justice and injustice." What we find, however, is Hume launching straightaway into a very complex argument for the claim that justice is an artificial virtue. Hume says that the artificial virtues are those that "produce pleasure and approbation by means of an

---

[8] Brown 1988, 76.

[9] Coleman 1992, 332. The meaning Coleman's use of 'internal connection' is unclear.

[10] Schauber 1999, 344. Schauber, however, does not provide an interpretation of the Motivation Argument. She bases her requirement of the necessary connection on what I think is a problematic reading of the argument(s) at T 3.1.1.22, SBN 465.

[11] There is another possibility for the Motivation Argument that is worth mentioning to avoid being charged with a false dilemma in presuming that the argument shows either a commitment to internalism or externalism. The thought is that the Motivation Argument can succeed on grounds that Hume himself need not and does not accept, and so the Motivation Argument is meant as a *reductio ad absurdum* of those moral theories that think morality is a system of rules and prescriptions that influence behavior. See Baier, 1991, p.184-5. See Radcliffe 1996, 391 for a discussion of this view both in Baier and Darwall. I do not think this response is successful given the apparent sincerity with which Hume makes claims about the motivational influence of morality.

artifice or contrivance, which arises from the circumstances and necessities of mankind" (T 3.2.1.1, SBN 477). The moral approbation of the artificial virtues depends on the establishment of human conventions regarding, for instance, the non-interference with the property and possessions of others. Hume's account of the origin of justice is quite sophisticated and I will have a bit more to say about it in chapter 7. My initial aim here is to point to an important argument that figures very prominently in Hume's account and creates problems for an internalist reading.

Hume argues that justice is an artificial virtue by establishing an "undoubted maxim," namely "*that no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality*" (T 3.2.1.7, SBN 479). Hume arrives at this *Undoubted Maxim* through an argument that is sometimes called *Hume's Circle*.[12] The circle presupposes that the moral worth of an action depends on the moral worth of motives and characters from which that action proceeds. For Hume, external actions themselves can be called virtuous or vicious, good or evil only in a derivative sense. We ordinarily take "actions as signs or indications of certain principles in the mind and temper" (T 3.2.1.2, SBN 477). In the previous chapter I gave some reasons for thinking that this claim is a part of Hume's theory of action and the will. Hume adds some further support. He says that when we blame anyone for failing *to perform* a "required" action, that we do so because they were not influenced by the "proper motive of that action" (T 3.2.1.3, SBN 477). Moreover, if we learn that the failure to perform in this case was due to external restrictions outside of the agent's control, but that she had the proper motive to act, we hold her in the same moral regard as if she had in fact acted.

With this premise in place, Hume can now argue for the Undoubted Maxim via Hume's Circle:

---

[12] The Circle is actually Hume's argument that justice, or a respect for the property of others, is an artificial virtue. He uses the Circle to establish the Undoubted Maxim, and then argues that if justice were a natural virtue, it would be a violation of the Maxim and nature would have established a "sophistry." So, we should conclude that justice is not a natural virtue. It is less than clear how this argument is supposed to work, and I will have more to say about it in chapter 7.

> To suppose, that the mere regard to the virtue of the action may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle. Before we can have such a regard, the action must be really virtuous; and this virtue must be deriv'd from some virtuous motive: And consequently the virtuous motive must be different from the regard to the virtue of the action. A virtuous motive is requisite to render an action virtuous. An action must be virtuous, before we can have a regard to its virtue. Some virtuous motive, therefore, must be antecedent to that regard. (T 3.2.1.4, SBN 478)

If S's doing Φ is virtuous, then the virtue of Φ stems not from S's external performance. The virtue of S's doing Φ stems from S's having some motive to Φ which, when contemplated by disinterested observers, gives rise in them to a unique pleasure. So, the "regard to virtue" is just the apprehension of such a motive. It is this motive that explains why the action is virtuous and so bestows upon it its merit. Apprehending this motive is not what *explains* the moral worth of an action since any explanation that appeals to the regard to virtue *alone* would be viciously circular.[13]

There is much more that can be said about this important and challenging argument. For now, I want to focus on Hume's attempt to forestall an objection to the Undoubted Maxim. Hume imagines that his interlocutors will point out that there certainly do seem to be cases where people are moved to act merely because they judge that they are morally obligated to do so. These seem to be cases where someone acts from a regard to the virtue of the act alone. Hume argues that the Undoubted Maxim does not rule out the possibility that people are sometimes motivated to act from a sense of duty alone. Given the importance of the *Sense of Duty Passage* to my argument, it will be helpful to quote it at length:

> When any virtuous motive or principle is common in human nature, a person, who feels his heart devoid of that motive, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it. A man that really feels no gratitude in his temper, is still pleas'd to perform grateful actions, and thinks he has, by that means, fulfill'd his duty. Actions are at first only consider'd as signs of motives: But 'tis usual, in this case, as in all others, to fix our attention on the signs, and neglect, in some measure, the thing signify'd. But tho', on some

[13] As Brown notes, Ross makes the same argument in *The Right and The Good* 1930, 4-6.

> occasions, a person may perform an action merely out of regard to its moral obligation, yet still this supposes in human nature some distinct principles, which are capable of producing the action, and whose moral beauty renders the action meritorious. (T 3.2.1.8, SBN 479)

The claim that someone can be motivated to act from a "certain sense of duty" or "regard to its moral obligation" is consistent with the conclusion of Hume's Circle. But how is this supposed to work?

Part of what Hume is saying is that motivation by the sense of duty alone depends on a certain degree of self- and other-awareness. The sense of duty is a response to an agent's realization that he is motivationally distinct from others. Take a variation on one of Hume's own examples: suppose the neglectful father recognizes his own "want of natural affection, which is the duty of every parent" (T 3.2.1.5, SBN 478). Hume takes this lack of natural affection to typically lead to neglect in the care of the child, and let us suppose that this neglect is serious but does not place the life of the child in peril. Now additionally suppose, as Hume evidently does, that this natural affection of parents for their children is one of the virtuous motives or principles "common in human nature" that he refers to in the Sense of Duty Passage.[14] The awareness that this is a common motive in human nature, together with the awareness that one lacks this motive, provide the conditions under which one can be motivated to act from a regard to one's duty and moral obligation. This seems to suggest that the father recognizes that it is his duty or obligation to provide care for his child, and that this recognition provides him with a motive to do so.

What makes it the case that it is the father's duty to provide care for his child? Hume uses the terms *duty* and *obligation* interchangeably in the Sense of Duty Passage. There are other places in the same section of the *Treatise* where he uses the locution "duty and obligation" and these cases seem to suggest that he treats the terms synonymously.[15] We can begin to make sense of the claim that the father has a duty to provide care for his child by looking at Hume's

---

[14] "Thus the relation of blood produces the strongest tie the mind is capable of in the love of parents to their children, and a lesser degree of the same affection, as the relation lessens" (T 2.2.4.2, SBN 352).

[15] T 3.2.1.9; and T 3.2.1.11.

account of obligation. There is good reason to think that Hume follows Hutcheson in distinguishing natural obligation from moral obligation. After he describes the process by which property and rules of justice become established though social convention, Hume says, "[t]he *natural* obligation to justice, *viz*. interest, has been fully explain'd; but as to the *moral* obligation, or the sentiment of right and wrong" Hume says that he will need to say more about natural virtues in order to speak of this latter sense of obligation (T 3.2.2.23, SBN 498). Natural obligation, for Hume as for Hutcheson, is a motive of self-interest. Moral obligation on the other hand presupposes a moral sense: "All morality depends on our sentiments; and when any action, or quality of the mind, pleases us *after a certain manner*, we say it is virtuous; and when the neglect, or non-performance of it, displeases us *after a like manner*, we say that we lie under an obligation to perform it" (T 3.2.5.4, SBN 517). We have a moral obligation to perform those actions the omission of which would give rise in disinterested and sympathetic spectators to an unpleasant feeling of disapprobation.

In our variation of the negligent father case we supposed the father to be aware both that parental affection is common to human nature and that he himself does not have this affection. This does seem to be a case where the non-performance of an action gives rise to disapprobation: "We blame a father for neglecting his child. Why? because it shows a want of natural affection, which is the duty of every parent. Were not natural affection a duty, the care of children cou'd not be a duty; and 'twere impossible we cou'd have the duty in our eye in the attention we give to our offspring" (T 3.2.1.5, SBN 478). To call natural affection and the care of one's children a duty is to signal that the lack of such affections and actions displeases after a certain manner, viz. displeases impartial and sympathetic spectators.

Internalism in the Sense of Duty Passage

The Sense of Duty Passage, then, seems to present the following picture: when an agent has self- and other-awareness enough to recognize that she lacks a common, naturally virtuous motive, she may judge that she is morally obligated to act as if she had that motive and so to act

as the naturally virtuous person would act. She is then motivated to act from her regard to this moral obligation or from her sense of duty understood as an apprehension that she is under a moral obligation. It is not clear whether the Sense of Duty passage demonstrates Hume's commitment to internalism, but I argue that there is some reason to favor an internalist reading of that passage. The most obvious difficulty that we're faced with in making this interpretation is that Hume does not need to accept a version of internalism in order for the Sense of Duty Passage to accomplish what Hume needs it to accomplish. Recall that the Sense of Duty Passage is offered as an attempt to forestall an objection to the Undoubted Maxim. The objection is that people do seem to be motivated to act from a sense of duty alone. It is possible that people may be motivated to do what they apprehend to be their moral duty without it being true that there is a necessary connection between such an apprehension and their motivational states. Indeed, Hume only says that "a person, who feels his heart devoid of that principle, *may* hate himself upon that account, and *may* perform the action without that motive, from a certain sense of duty" (T 3.2.1.8, SBN 479; emphasis added).[16] Additionally, an internalist reading of the Sense of Duty Passage does not seem to explain why Hume appeals to the agent's self-hate.

In what follows, I argue that Hume thinks there are some judgments, namely judgments about our own moral obligations, which are themselves motivational states. There are two premises to the argument. The first is that the feeling of hatred to which Hume refers in the Sense of Duty Passage is the moral sentiment of disapprobation directed at oneself. Moral self-disapprobation just is self-hate. The second premise of the argument is that self-hate is itself a motivational state. There are a number of difficulties that this interpretation faces, and I will conclude this section by responding to them. I argue that we have reason to favor an internalist reading, even though there is an externalist reading that is equally textually well-grounded.

---

[16] The second point, that someone may act without the motive, is consistent with weak versions of motive internalism and so I will not have anything else to say about it here. The second point, that someone may hate himself upon reflection, will be dealt with when I ask whether Hume thinks we are morally obligated to take up the common point of view.

Moral Self-disapprobation is Self-hate

Radcliffe argues for the claim that self-hate is a deliverance of the moral sense by relying on some problematic textual support. While I think that Radcliffe's identification of self-hate as moral self-disapprobation is correct, I do not think her textual evidence is particularly strong. First, she claims that when Hume distinguishes between the calm and violent passions, he lists the moral sentiments as well as the sense of beauty as examples of the calm passions.[17] What he says is, "of the first kind [calm passions] is the sense of *beauty and deformity in action*, composition, and external objects" (T 2.1.1.2, SBN 276; emphasis added). Radcliffe must think that by the "sense of beauty and deformity in action" Hume is referring to the moral sense. Pointing to places where Hume uses the expression "moral beauty" might help this interpretation.[18] Still, there are only a handful of places throughout the *Treatise* where Hume uses the phrase "sense of beauty" and this is the only passage where he refers to the sense of beauty in actions. In one place where Hume later describes the sense of beauty, he talks about it arising through principles of sympathy in a manner that is very similar to the way in which sympathy gives rise to moral evaluations. He says, for instance, "the same principle [sympathy] produces, in many instances, our sentiments of morals, *as well as* those of beauty" (T 3.3.1.9, SBN 577; emphasis added). This language would be a bit surprising if, as Radcliffe suggests, the sense of beauty in action just is the moral sense. I do not think this objection is decisive against Radcliffe. Her rendering of Hume's singular occurrence of the "sense of beauty in action" phrase may be the best available. However, it is not obviously so and deserves at least some comment.

Suppose we accept that Hume was referring to the moral sense in his initial list of the calm passions, how does this show that the moral sentiments are motivational? Radcliffe mentions Hume's later discussion of the calm passions, where he notices that these "calm desires and tendencies" are divided into two kinds, "either certain instincts originally implanted in our

---

[17] Radcliffe maintains this reading in 2006, 357.

[18] For instance, T 3.2.1.8.

natures, such as benevolence and resentment, the love of life, and kindness to children; or the general appetite to good, and aversion to evil, consider'd merely as such" (T 2.3.3.8, SBN 417). Since we are supposing that Hume has already told us that the moral sentiments are calm passions, since he now says that some of these calm passions are calm desires and tendencies towards good and away from evil, Radcliffe thinks there is good reason to think that the "general appetite to good, and aversion to evil, consider'd merely as such" is identical to the moral sense. The moral sentiments, then, would themselves be desires and aversions, and their motivational role would be secured. But this reading is tendentious. Hume does not equate good and evil with *moral* goodness and evil. In fact, Hume often uses locutions like "good and evil, or pleasure and pain," which suggests that a more natural reading of this passage refers to a calm desire for pleasure over pain. Radcliffe suggests this alternative rendering in a footnote. She rejects it, however, because it "ignores Hume's inclusion of the moral sense among the calm passions at T 276."[19] This makes the demand for rendering the "sense of beauty in action" passage as referring to the moral sense all the more pressing. That is Radcliffe's one and only link between the (perhaps) moral language of good and evil in general and the calm passions. Without it there doesn't seem to be any good textual reason to identify the general appetite to good and aversion to evil with the moral sense.

Perhaps Radcliffe did not mean to *identify* the moral sense with the general appetite for good and aversion to evil. This suggestion is given some plausibility by the fact that she makes a slight adjustment to her view in a later paper. There she says that, "there is good reason to think that the general appetite to good and aversion to evil are identical to *or include these senses* [the moral sense and the sense of beauty]."[20] The second half of the disjunction suggests the view that the calm sentiments of approbation are among the general appetites for good. As a particular kind of general appetite for good, the sentiment of approbation need not share all its properties

---

[19] Radcliffe 1996, 397-8, n.30.

[20] Radcliffe 2006, 357; emphasis added.

with other appetites for good. For instance, while other appetites for good might not be pleasurable states, moral approbation might be unique as a pleasurable kind of desire. I do not think this interpretation is plausible. This is because the claim that the moral sentiments are among, and not identical to, the general appetite for good and aversion to evil still depends on the identification of the "sense of beauty in action" with the moral sense, and so it too will succumb to the same textual criticism I offered above.

These two passages seem to be Radcliffe's main premises in her argument that moral approbation and disapprobation are motivational passions. She does claim that Hume's discussion of the virtue he calls strength of mind corroborates her reading, but it cannot provide independent grounds for adopting that reading.[21] While I agree with Radcliffe's conclusion, that moral approbation and disapprobation are motivational passions, I do not accept her textual support. I will conclude this section by presenting some considerations in favor of identifying self-hate with a deliverance of the moral sense directed at oneself. In the subsequent section, I will continue to support Radcliffe's conclusion by attempting to show that self-hate is motivational.

Recall that Hume thinks the terms 'duty' and 'obligation' are interchangeable. When Hume later turns his attention to the obligation of keeping one's promises, which derives from the artificial virtue of fidelity, Hume again claims that the "sense of duty supposes an antecedent obligation" (T 3.2.5.6, SBN 518). So, it must be the case that if a person is moved to act in a particular way from a sense of duty, then there is some antecedent obligation in this case, and Hume makes it clear that this obligation is an obligation to perform those actions that the naturally virtuous person would or an obligation to acquire the naturally virtuous person's quality of mind. The natural antecedent obligations are the natural motives of beneficence, limited generosity, affection (and etc.). I think it is this presupposition of antecedent *obligation* in the Sense of Duty Passage that allows us to identify the feeling of self-hate with moral

---

[21] See Radcliffe 1996, 398; and 2006, 357.

disapprobation. The negligent father "hates" himself because he lacks a common virtuous motive. This lack gives rise in disinterested observers to the unique uneasiness of moral disapproval, and therefore having natural affection is *morally* obligatory. The apprehension of his lack of natural affection gives rise in both the father and the moral spectator to an uneasy sentiment, "hate" in the former case and moral disapprobation in the latter. I am arguing that Hume thinks that, if we're using the negligent father example to understand the Sense of Duty Passage, the negligent father occupies the same sort of disinterested, moral point of view as does the moral spectator.

The Sense of Duty Passage presents Hume's response to the question "may not the sense of morality or duty produce an action, without any other motive?" (T 3.2.1.8, SBN 479). The crucial question is what Hume means by the "sense of duty." If we understand Hume to mean by the "sense of duty" something like a belief that other people call certain actions and motives "moral duties," then he would not be answering *this* question. The belief that others call something a "duty" cannot produce an action without presupposing another motive, such as a desire for the praise that one thinks the so-called "dutiful" person acquires. I do not think Hume takes the "sense of duty" to be a belief about what others call a "duty." Instead, I think he takes the "sense of duty" to be an apprehension of what is actually a duty. I am putting forward the suggestion that apprehending what is actually one's duty requires that one, at some point, engage in contemplation of the relevant ideas from a disinterested and sympathetic point of view. This, together with the fact that the negligent father feels a negative sentiment of the same kind as the moral spectator (i.e. painful and uneasy), lends some plausibility to the suggestion that self-hate is a moral sentiment.[22]

---

[22] The argument here is not as compelling as I would like. I admit there is some plausibility to reading the Sense of Duty Passage as illustrating cases where one is moved to act, not via motivating moral sentiments, but via desires additional to one's recognition that she is different from others. My suggestion here is just that this sort of a reading doesn't seem to explain the sense in which the Sense of Duty Passage illustrates cases of agents being moved by considerations of *duty* or *obligation.* Also, as I argue below, the reading I favor is aided by the difficulty in identifying other plausible candidates to identify with the "hate" the negligent father feels.

Self-hate is Not Genuine Hatred

I do not think that the foregoing considerations are decisive. One question that immediately arises is why Hume chooses to employ the language of hatred when he could have used approbation and disapprobation instead. It may be the case that Hume does not want us to read the Sense of Duty Passage as resulting in a moral sentiment, but that he thinks that motivation from the sense of duty stems from some other sort of sentimental response. In the next section I will examine the possibility that self-hate is identical to genuine hatred, and I will reject this suggestion. [23] Then I will examine the possibility that self-hate is identical to humility (roughly, Hume's term for shame, or the contrary of pride). I think it is much less clear that this isn't a plausible alternative, so will only tentatively dismiss it. The arguments of the following sections, then, together provide something like an argument from elimination: since self-hate is neither hatred nor humility, there is more reason in favor of accepting the identification of self-hate with moral self-disapprobation.

When I refer to *hatred* or *genuine hatred* I mean to be referring to the indirect passion that Hume describes in T 2.2. Hatred is the contrary of love (a passion that Hume understands very broadly to include esteem, respect, compassion, and romantic and sexual affection).[24] To say that hatred is the contrary of love I mean that both love and hatred have the same kinds of objects but have different causes insofar as the painful passion of hatred is caused by other painful perceptions, while the pleasurable passion of love is caused by other pleasurable perceptions.[25] Nevertheless, Hume makes it clear that both love and hatred have the same kind

---

[23] While she does not make this explicit, I think that Coleman's argument that Hume accepted internalism presupposes that self-hate is a genuine species of hatred. But even if she were to reject this claim and hold that self-hate is moral self-disapprobation, she still has to maintain that self-hate functions in the same way that hatred does (Coleman 1998, 335). It is not clear how her view counts as a version of internalism as I understand it here. After all, the externalist is perfectly happy to say that moral judgments sometimes cause desires to act accordingly.

[24] Cohon 1997, 269.

[25] When you deliberately inflict wanton injury on me, I have an idea of you that I associate with you as the cause of pain. My feeling of pain resembles the painful feeling of hatred. There is, then, both the

of object, and it is this commitment of his moral psychology that seems to rule out the possibility that self-hate is a genuine species of hatred. Hume says that

> love and hatred are always, directed to some sensible being, external to us, and when we talk of *self-love*, 'tis not in a proper sense, nor has the sensation it produces any thing in common with that tender emotion, which is excited by a friend or mistress. 'Tis the same case with hatred. We may be mortify'd by our own faults and follies; but never feel any anger or hatred, except from the injuries of others. (T 2.2.1.2, SBN 329-30)

One way of reading this last sentence presents a problem. Hume says that we may be mortified by our own faults but not feel hatred. That we *may* feel this way without feeling hatred suggests that there are times where we may feel this way and feel hatred towards ourselves. If hatred is merely a feeling, a view suggested by this passage, then there doesn't seem to be any reason why I can't feel hatred towards myself. However, Hume's theory of the indirect passions crucially depends on one's not being able to genuinely hate herself. This is because the objects of love and hatred are what distinguishes them from the related indirect passions of pride and humility. The causes of pride and love are the same (other pleasurable perceptions) and the causes of humility and hatred are the same (other painful perceptions).[26] As we'll see presently, I think part of Hume's thinking here is that pride/love and humility/hate are phenomenologically rather similar in feeling. So, Hume wants to say, pride and love are distinguished, as humility and hatred are, by their *objects* (T 2.2.2.3, SBN 333). If I may feel hatred towards myself then it is no longer clear how we are to distinguish humility from hatred.

As I mentioned above, there does seem to be tension in Hume's view on the passions. On the one hand, the indirect passions are accounted for by Hume's ingenious application of his principles of association according to his doctrine of the double relation of impressions and ideas. This suggests that the passions are to be distinguished by their casual origins. However,

---

association of ideas and the association of impressions that Hume thinks is causally responsible for the passion of hatred. This is the double relation account that I described above.

[26] Strictly speaking, these are only the *qualities* of the causes of these passions that we attribute to the *subjects* of these causes.

Hume also suggests that the passions are feelings and can be distinguished on the basis of feelings alone. While I cannot propose an adequate solution to this problem here, I do think a few remarks might help us to see what Hume is up to with the hatred/self-hate distinction. Hume recognizes that many perceptions are so subtle that we become aware of them more by their effects than by direct introspection, and this is especially so for those of us that do not reflect on our own states with the same care which Hume demands: "'Tis natural for one, that does not examine objects with a strict philosophic eye, to imagine, that those actions of the mind are entirely the same, which produce not a different sensation, and are not immediately distinguishable to the feeling and perception" (T 2.3.3.8, SBN 417). When I am mortified at my own faults and follies, I may not be able to tell what exactly I am feeling about myself. I may easily mistake this feeling for hatred when I introspect on those passions that arise when I recognize that I have been remiss in my duties. I think that Hume's point here speaks to something that is fundamentally correct, especially for those individuals who are not used to reflecting on their faults, perhaps because of their immaturity. Teenagers are apt to say that they hate themselves when they come to learn that they are deficient in some way (though, of course, their immaturity also leads them to judge more deficiencies than actually exist). Due to their inexperience in differentiating their emotional states they are apt to confound the displeasures of coming to perceive a moral failing in themselves with a feeling of hatred directed at themselves. With more reflective experience we can come to make finer-grained distinctions between our experientially similar passions. [27]

Moreover, Hume insists that hatred and self-hate are different because they do not typically occupy the same sort of causal role. Hume requires that different effects have different

---

[27] I am here suggesting that there is a sense in which hatred and humility, along with love and pride, do not produce different sensations and a sense in which they are not immediately distinguishable. I think the idea here has to do with the fact that each pair of passions shares the same kind of cause (considered as a quality not a subject). Hatred and humility are both caused by painful perceptions and are themselves painful perceptions, and it is in virtue of this that these passions are apt to be confounded by someone that lacks experience in introspection.

causes, so if hatred and self-hate give rise to different states of affairs, then there is some good reason for attributing this difference in effect to a difference in cause. Now, of course, Hume says that "passions may express themselves in a hundred ways" (T 2.2.6.5, SBN 368). He thinks that the relationship between a passion and its typical effects is "abstractedly consider'd…not necessary," and so he holds that there is no contradiction in supposing that "love might have had the same effect as hatred, and hatred as love" (T 2.2.6.6, SBN 368). Nevertheless, given the way that human beings are in fact constituted, Hume thinks that there are regular, universal, causal connections between certain passions: "The passions of love and hatred are always follow'd by or rather conjoin'd with benevolence and anger… hatred produces a desire of the misery and an aversion to the happiness of the person hated" (T 2.2.6.3, SBN 367). Now, if Hume meant to identify self-hate with hatred, then when we turn to the Sense of Duty Passage we should expect to find the self-hating agent experiencing a desire for her own misery and an aversion to her own happiness. However, what Hume says is that the person who has no natural gratitude may nevertheless still feel pleased by performing grateful actions. The Sense of Duty Passage is supposed to illustrate the possibility of an agent being moved to act from a regard to the virtue of her action alone. If she were aware that doing as the generous person does would cause her to feel pleasure, and she genuinely hated herself for lacking natural generosity, then it seems her hatred would undermine her motivation to act since it would produce in her an aversion to her own happiness. Reading self-hate as genuine hatred would not allow Hume to explain the phenomenon he set out to explain in the Sense of Duty Passage. It is important, then, for Hume to distinguish genuine hatred from cases of self-hate.[28]

---

[28] Hume does not seem to think that self-hate is one of the principal causes of the miseries that lead people to suicide. In "Of Suicide" ([1755] 2008) Hume says that people are driven to suicide primarily from "age, sickness, or misfortune [that] may render a life a burden, and make it worse even than annihilation," but he recognizes that others may be driven to suicide by an "incurable depravity or gloominess of temper" (323-4). In this essay, Hume makes one reference to a "hated life" (320). If we interpret *hate* here as an instance of genuine hatred, then we would expect the person to feel anger towards his life and so have more motive to prolong his suffering. The hatred of one's life, like the hatred of oneself, should be understood as genuine hatred.

Self-hate as Humility: An Argument for Externalism

Since self-hate does not seem to be the same sort of thing as genuine hatred, perhaps Hume means to identify self-hate as some other kind of indirect passion. Humility is a seemingly plausible candidate since it, like self-hate, takes oneself as its object. Charlotte Brown seems to adopt this view. She claims that the person in the Sense of Duty Passage that lacks the common motives feels self-hate and thereby feels ashamed, which is what she calls humility. Since a sense of self-worth is an important ingredient to our happiness, we lose out on something important to our happiness when we lack certain natural motives. So pride together with the desire to be happy give rise to the sense of duty when one's own moral sense disapproves of one's lack of common motives. The disapproval of the moral sense is motivationally relevant to people because they desire to have character traits of which they are proud as a part of their happiness.[29]

The sense of duty works by motivating through a regard to the moral worth of the action, however, as Brown understands the view, the motivation only arises by connecting shame (self-hate) with a desire to be happy: "It is not the regard for the moral worth of the action by itself which prompts a person to do what is right."[30] This is a version of externalism that Brown calls the "trigger theory," since regard to the moral worth of the action is not what provides motivation.[31] Moral self-disapprobation motivates because it is shameful and because pride in her own character is a part of the happiness that she already desires for herself. The shame triggers her desire to be happy so that she is thereby moved to act in a way she would not be ashamed of. Moral approval acts in a supporting role to one's desire for pride as a part of one's happiness, but it does not itself factor as a motive to action. So, while Brown maintains that

---

[29] Brown 1988, 82.

[30] Ibid., 83.

[31] Brown externalist reading is similar position to that is developed by Jonathan Harrison. See Brown 1988, 73; Harrison 1976, 5-15.

Hume is committed to a version of internalism in the Motivation Argument, she thinks that the Sense of Duty Passage shows that Hume did not develop the internalist account he required. So, Hume is inconsistent.

As I mentioned above, a view like Brown's that seeks to identify self-hate with humility does have the advantage that humility, unlike hatred, takes the self as its object. However, humility and pride, unlike hatred and love, are not causally related to motivating passions so it's not clear how this account could explain the motivational role of self-hate in the Sense of Duty Passage.[32] Brown, of course, is aware that humility is not motivating in the way that love and hatred are. So it might seem that she is being inconsistent when she claims that shame causes a desire for happiness.[33] However, Brown does not think that shame is the cause of the desire for happiness. She points out passages from the second *Enquiry* where Hume seems to say that pride is an ingredient in one's happiness. He says that being virtuous is "attended with a pleasing consciousness or remembrance," and that "[i]nward peace of mind, consciousness of integrity, a satisfactory review of our own conduct – these are circumstances very requisite to our happiness" (EPM 9.2.23, SBN 283). Brown is not committed to an inconsistency because Hume holds that pride in one's own character is a feature of one's own happiness, and one's own happiness is something that one desires independently of ever having felt ashamed of one's lack of common virtuous motives.

Brown's reading is compelling and does seem to fit well with certain portions of Hume's theory. For example, Hume appeals to various typical causes of pride and humility in order to test his theory of the passions, that is, in order to confirm that these passions depend on the double relation of impressions and ideas. He argues that, no matter if one is a Hobbesian, a sentimentalist, or a rationalist, it is still the case that the perception of oneself as being virtuous

---

[32] See my discussion in chapter 5 of the "Weak and Strong Passions." Note, too, that there is some connection between pride/humility and motives. I will have more to say about this connection below.

[33] Stefan Kalt 2005, 150 presents this objection.

gives rise to pleasure, and the perception of oneself as vicious gives rise to pain. Given that in either case it is *my* virtue or *my* vice that is contemplated, the association of ideas takes place since I associate my idea of virtue and vice with my idea of myself. Given that the pleasurable feeling of virtue resembles the pleasure of pride, and the painful feeling of vice resembles the pain of humility, the association of impressions also occurs. Hume says that for the sentimentalist "the pain and pleasure, therefore, being the primary causes of vice and virtue, must also be the causes of all their effects, and consequently of pride and humility, which are the unavoidable attendants of that distinction" (T 2.1.7.5, SBN 296). In Book 3, when Hume is considering the unique pleasure and pain that are involved in moral evaluation, he briefly reminds us of the double relation account of the indirect passions and says:

> Now virtue and vice are attended with these circumstances [those required by the double relation account]. They must necessarily be plac'd either in ourselves or others, and excite either pleasure or uneasiness; and therefore must give rise to one of these four passions; which clearly distinguishes them from the pleasure and pain arising from inanimate objects, that often bear no relation to us: And this is, perhaps, the most considerable effect that virtue and vice have upon the human mind. (T 3.1.2.5, SBN 473)

These passages suggest that the apprehension of oneself as virtuous causes pride while the apprehension of another's virtue causes love. The apprehension of oneself as vicious causes humility while the apprehension of another's vice causes hatred. There is then some further textual support for Brown's interpretation of the Sense of Duty Passage, which holds that the moral sense disapproves of one's lacking common virtuous motives and so causes self-hate, understood as humility, which, because of its connection with happiness, gives rise in one to a desire to act as the virtuous person would. This consideration favors an externalist interpretation because it holds that the motivation is caused by a desire for happiness, not by the moral sentiment itself. The moral sentiment simply informs the agent that she will experience shame-producing disapprobation unless she acts as the virtuous person would, and given her desire to avoid feeling shame she is moved to do so. Supposing that she had no desire for her own

happiness, and so no aversion to feeling ashamed, her judgment that she would feel ashamed of herself for not acting as the virtuous person does would not move her to act.[34]

An Internalist Rejoinder

Brown's position still leaves some questions unaddressed. First, and most obviously, is why Hume failed to take note that his own theory differed significantly from the theory he requires for the success of the Motivation Argument. This inconsistency arises between two rather important arguments that occur very near each other in the text, so it's hard to understand how Hume could have overlooked it. Another concern with the interpretation is that it still doesn't quite address why Hume talks about someone coming to hate herself on account of her lacking common motives. If my desire for my own happiness is what serves as my motive and avoiding painful perceptions will help me to be happy, all I need is to recognize that my failing my obligations will make me unhappy and the motivation is supplied. Why does Hume need to bring in a discussion of self-hate, understood as humility, at all? I think that these two concerns warrant looking at an internalist interpretation of the Sense of Duty Passage.

If Hume accepted internalism this would have amounted to a rejection of Hutcheson's position. Recall that for Hutcheson the moral sentiments are not themselves motivational. He sharply distinguishes justifying from exciting reasons so that the deliverances of the moral sense do not themselves provide motives for action. Our motivation stems from our natural capacities alone and we would be in full possession of these dispositions even if it were the case that we had no moral sense at all. Similarly, we might be in full possession of a moral sense, but without our natural benevolent dispositions, there would be no motive for us to pursue the well being of

---

[34] On the internalist reading, however, she would still have some motive to act. The passage from T 3.1.2.5 says that the moral sentiments cause either love and hatred or pride and humility. On the internalist reading of the Sense of Duty Passage, even though self-hate is not identical to humility, it still gives rise to humility in addition to the motivating moral sentiment. For the person who desires her own happiness, the painful perception of shame can serve to strengthen the motivating moral sentiment to act as the virtuous person does. The person who has no desire for her own happiness would still be moved to act by the moral sense, but this motive would not be reinforced by her consideration of the pain of humility.

others for its own sake. In the previous chapter we saw that, for Hutcheson, human nature was created so that there is in fact a regular connection between our motivational states and our states of approval, and that this result is the deliverance of a divine benefactor. Hume would not accept this last claim. And it seems too that Hume does not accept Hutcheson's distinction between exciting and justifying reasons for action.

The crucial question that needs to be addressed here is where the motivation to act according to the sense of duty comes from. If the motivation arises from some source independently from the sense of duty itself, then the apprehension that some course of action is one's duty is not directly motivational and Hume would not accept internalism. This is Brown's interpretation and, as we've seen, it doesn't seem to explain the role of self-hate. Brown identifies self-hate as humility, but this only adds to the mystery. Humility is not itself a motivational state, nor is it connected with desires in the same regular and universal sort of way that love and hatred are.

Self-hatred, regardless if it's a sort of humility or a moral sentiment, is presumably painful. There are two main options for understanding the motivational influence of painful states. The first is that pain serves merely as the object of aversions that agents already possess. On this view, agents have a pre-existing desire for pleasure and a pre-existing aversion to pain. Any particular pleasure or pain, then, is related to an agent's motivational states only by "triggering" this desire. The other view maintains that pleasure and pain motivate directly, without any pre-existing desires or aversions. If this latter view is Hume's, then self-hatred as a kind of pain may be directly motivational.

Radcliffe discusses Locke and Hobbes on pleasure and pain to show that there is some historical precedent for maintaining that pleasure and pain are directly motivational. She thinks that Hume held the view that "to be in a painful state is to be in a state one has a motive to be relieved of, to be in a motivational state without the presence of previous desires."[35] The main

---

[35] Radcliffe 1996, 400.

support for this interpretation is Hume's initial discussion in Book 1 of how desire and aversion usually arise.[36] Hume says that pleasure and pain are copied as ideas, and then when they return "upon the soul" they produce new impressions of reflection, desire and aversion. As Radcliffe puts it, ideas of pleasure and pain *become* desires and aversions on this model.[37] Moreover, Hume does say that the passions arise either directly or indirectly from pleasure or pain, and he says that the "chief spring or actuating principle of the human mind is pleasure or pain" (T 3.3.1.2, SBN 574). So, for Radcliffe, all of the direct passions are motivational since they all proceed from pleasure and pain directly.[38]

So, how does the sense of duty motivate for Radcliffe? First, when a person lacks a virtuous motive common to human nature and that person imagines her own actions from the perspective of an impartial observer she experiences self-hatred. This self-hatred is a moral sentiment that Radcliffe calls self-disapprobation. This is a painful sentiment and it motivates the agent directly to perform those actions that a naturally virtuous person would perform. The painful self-disapprobation is relieved only because we have habituated ourselves, in our observations of others, to take actions as the signs of character. Radcliffe says, "the motivation to do an act a virtuous person would do naturally derives from the moral sentiment which is the source of moral justification. Thus, this is an internalist account of moral motivation."[39]

There are two main reasons for rejecting Radcliffe's interpretation, neither of which is entirely successful. The first is that, if we accept that self-hate is moral self-disapprobation and that the moral sentiments are motivational, then Hume would hold that the moral sentiments are direct passions. Radcliffe is clear that this is a commitment of her view. However, it is not clear

---

[36] See my chapter 5 discussion of "The Multiple Divisions of the Passions."

[37] Ibid.

[38] Though we'll see shortly that the view is more complicated since Hume acknowledges the existence of productive direct passions.

[39] Ibid., 402.

how this could be the case. Hume says that the direct passions "arise immediately from good or evil, from pain or pleasure" (T 2.1.1.4, SBN 276), but the moral sentiments do not arise in this way. This is because the moral sentiments arise from a disinterested perspective through the mechanism of sympathy.[40]

Radcliffe may attempt to respond to the objection as follows: The objection rests on a mistaken conception of the direct passions. It is true that Hume does introduce the distinction between the direct and the indirect passions in such a way so as to make it appear as if it were his view of the direct passions that they *only* stem immediately from pleasure and pain. However, there is good reason to think that Hume's view on the matter is more complicated. For example, Hume thinks that desires are direct passions and that love and hatred are indirect passions. As we have seen, Hume maintains that there is a very strong causal connection between love and benevolence in human nature. Since benevolence is a desire for the happiness of another, Hume is committed to the view that some direct passions have their causal origin in the psychological mechanisms of the indirect passions. Given the complicated causal origin of benevolence, then, it doesn't seem that the mere fact that some passion has a complex causal origin is sufficient proof that the passion in question is not direct.[41] So, the first objection to Radcliffe is unsuccessful.

The second reason to reject Radcliffe's interpretation has to do with general considerations about the relationship between sentiments and motives on Hume's view. While Radcliffe recognizes that there are important differences between these, the view that pleasure

---

[40] Schauber and Kalt both identify this problem, but they think it is a problem for seemingly different reasons. Schauber claims that the moral sentiments cannot be direct passions because the "principle of sympathy does not operate within us automatically, but requires an effort (or at least moral training)" (1999, 349). Kalt, however, describes sympathy as a "mechanical process," but still rejects reading the moral sentiments as direct passions on the grounds that "the ascent to the general point of view at which sympathy occurs is not mechanical, but implies a conscious, even practiced, effort on the part of the sympathizer" (2005, 147).

[41] See McIntyre 2000 who argues that Hume is committed to the view that the direct passions always arise in the context of the indirect passions. While this view is interesting, it need not be defended here. Radcliffe can respond to Schauber's and Kalt's objection from sympathy by showing that *at least some* direct passions do arise in a more complex manner than they suppose.

and pain are themselves motivational seems to threaten this distinction. If self-disapprobation is motivational because it is a sort of pain, and pain motivates directly, then we should expect to find that other painful passions have motivational influence in virtue of the direct motivational capacity of pain. And the same should hold *mutatis mutandis* for pleasure. The cause of humility is always some particular pain just as the cause of pride is always some particular pleasure.[42] If pain and pleasure are themselves directly motivational, then humility and pride would seem to be motivational as well. However, Hume denies that pride and humility have a direct motivational capacity and so one wonders what it means to say that pain and pleasure are directly motivational when some painful and pleasurable passions are not motivational in this way.[43] It seems as if the account cannot be fully general.

I think that Radcliffe can respond to this criticism as well. Hume does not think that pride is entirely unrelated to desire. Recall Hume's discussion of one's pride in owning a "suit of fine cloaths" in the previous chapter. Hume thinks the indirect passions "being always agreeable or uneasy, give in their turn additional force to the direct passions, and encrease [*sic*] our desire or aversion to the object" (T 2.3.9.4, SBN 439). This seems to confirm Radcliffe's reading that pleasure and pain are directly motivational because Hume is saying that whatever motivational relevance pride and other indirect passions have stems from their being pleasurable or painful states: "the pleasure, which attends that passion [pride], returns back to the direct affections, and gives new force to our desire or volition, joy or hope" (ibid.). In order to avoid any inconsistencies, then, we need to make sense of how this view fits with Hume's claim that "pride and humility are pure emotions of the soul, unattended with any desire, and not immediately exciting us to action" (T 2.2.6.3, SBN 367). This claim is supposed to mark unique features of

---

[42] Strictly speaking, these are only the *qualities* of the causes of these passions that we attribute to the *subjects* of these causes.

[43] Kalt raises this objection to Radcliffe. Kalt's presentation of the objection relies heavily on his claim that 'motive' and 'desire' are synonymous for Hume. I have tried to remain neutral on this issue here, and I think the objection to Radcliffe can be presented without making that assumption.

love and hatred compared to pride and humility "which in so many other particulars correspond to each other" (ibid.). I think the idea here is that, whereas pride and humility have motivational influence on desires that are already present in the soul, love and hatred have the capacity to motivate by giving rise to new desires *in addition* to strengthening desires that already exist. Pride, unlike love, only operates on particular desires that are already directing the agent to act in particular ways. Therefore, there is no reason to think that pleasure and pain are completely motivationally inert with respect to pride and humility simply because Hume describes the latter as being "compleated within themselves" (T 2.2.6.3, SBN 367).

The idea here is that pleasure and pain are directly motivational, but not all pleasurable and painful states are motivational in the same way. Pleasure and pain have a motivational role in cases of pride and humility only when there exists a desire or aversion that shares a causal history with this particular occurrence of pride and humility. This happens when the desire is caused by the same pleasure that is the partial cause of the pride and when the aversion is caused by the same pain that is the partial cause of the humility. In these cases, the pleasure and pain of pride and humility serves a reinforcing and strengthening role. If I didn't desire the fine suit of clothes, then the feeling of pride, pleasurable as it may be, would not motivate me to buy the clothes. However, when I love or hate someone, the pleasure and the pain of these passions may give rise in me to a desire or aversion that I did not have before.[44] The motivational influence of love and hatred is not limited to reinforcing or strengthening present desires and aversions.

---

[44] I am not here intending to settle an interpretive dispute that involves understanding how Hume thinks one's hedonistic beliefs, beliefs that pleasure or pain for oneself are likely to arise from an object, give rise to motivational states. One option maintains that there exists, at all times, desires for pleasure and aversions to pain and one's hedonistic beliefs merely direct these background motivations. Another option holds that desires and aversions for pleasurable or painful *objects* arise under the influence of a lively hedonistic belief in each case. Cohon (2008, chapter 2) notes Radcliffe's attempt (1999) to describe a third alternative: motivation only arises with hedonistic beliefs given certain traits of character (being cowardly, having a sweet-tooth, etc.). No matter what model is selected for understanding Hume's view of the motivational influence of hedonistic beliefs, it will still be the case that the motivational role of love and hatred will be distinct from the motivational role of pride and humility.

Brown's Response

It is open to Brown to respond to these arguments by reminding us of where she locates the source of motivation in the Sense of Duty Passage. Again, on Brown's view, the moral disapprobation gives rise to self-hate or humility and this in turn triggers one's desire to avoid shame insofar as she has a desire for her own happiness. The motivation stems from the desire for happiness on her account, it does not stem from humility. The negligent father does not form a new desire to act as he ought, but he is moved to act by seeing that action as a means to satisfying his pre-existing desire for his own happiness. Humility triggers and reinforces his desire for his own happiness in the same way that pride triggers and reinforces one's desire to own the fine suit of clothes. Therefore, Brown's reading cannot be rejected by pointing to the motivational limitations of humility.

The Sense of Duty Passage does not provide a clear case of a necessary connection between one's judgments about her moral obligations and her motivational states. There are a few reasons why this passage can plausibly be read as illustrations both of internalism and externalism. One reason comes out in the Radcliffe and Brown exchange above: Hume recognizes that pleasure and pain are motivational in different ways. Sometimes pleasure and pain give rise to new desires and aversions, as when ideas of pleasure or pain return "upon the soul," but other times pleasure and pain reinforce desires and aversions that are already present and connected with pride and humility. So, the mere fact that pleasure and pain are connected with motivational states cannot settle the issue in favor of internalism, as Radcliffe supposes.

These last sections show that the Sense of Duty Passage can plausibly be read as offering versions of internalism and externalism. I do not think that the externalist reading is implausible. However, I think that we should read the Sense of Duty Passage as the internalist recommends. This is because, if we accept the externalist reading, then, as Brown observes, Hume fails to develop the internalist account he needs in order to support the Motivation Argument against the rationalist. Reading Hume as an externalist makes him inconsistent. Since there is equally good textual support for reading Hume as internalist, and reading him as an internalist makes him

consistent, I think the internalist reading is more plausible. I do not wish to defend the view that any plausible interpretation of a historical figure is plausible only if it renders that figure's view consistent. However, I do accept the following principle: if two (or more) competing interpretations are equally textually well-grounded, then that interpretation should be preferred that does not render it inconsistent with other elements of the theory. This principle indicates that we should prefer the internalist reading to the equally textually well-grounded externalist reading.

However, if we adopt an internalist interpretation of Hume then we find ourselves faced with a fairly large tension in his view. One might think that this tension is large enough to override the application of the interpretive principle I just offered: given this tension it is no longer the case, one might argue, that both internalism and externalism are equally textually well-grounded. The tension arises when we consider that the moral sentiments arise only from certain perspectives, namely the disinterested and sympathetic perspectives of what he calls the common, or general, point of view. It is not clear from the text that taking up such a perspective actually gives rise in observers to the relevant feelings, since it may be the case that all that is required for moral evaluation is that one imagine what she *would feel* were she to occupy such a perspective. If moral evaluation only depends on the hypothetical feelings of observers, then it is entirely unclear how Hume could maintain a version of internalism. I may perform a moral evaluation from the common point of view and since I only need to have certain hypothetical sentiments, which are not motivating since they aren't my actual sentiments, it seems my moral evaluation is divorced from my motivational states. I turn to this problem in the next section.

Internalism and the Common Point of View

Before delving into the conceptual tension that seems to arise between understanding moral evaluation arising from a common point of view and internalism, I think it will be helpful to specify the sort of internalism that we have so far located in Hume. The idea here is that the sense of duty motivates by giving rise in an agent to a motivating moral sentiment. Radcliffe

takes this view to be a version of appraiser internalism, but a particular version of appraiser internalism. Appraiser internalism for Radcliffe is the view that "if a person genuinely accepts or embraces certain moral judgments, she has some motivation to do what the judgment requires in the relevant situation."[45] Notice that this formulation itself does not settle the question about the necessity of the relationship between moral judgment and motivation, but Radcliffe does seem to accept such a relationship. She says that Hume's version of internalism recognizes that human psychology is contingent, and that humans could have been constituted differently than we are. However, "given our actual constitution, morality is *sufficient* for motivation."[46]

Radcliffe adds certain features to the version of appraiser internalism that she attributes to Hume. For instance, she says that this account of internalism makes some presuppositions about the relationship between moral and non-moral properties. Radcliffe says that it assumes that when some moral quality is attributed to Φ (that Φ is virtuous, wrong, etc.), then there are some non-moral properties of Φ that give Φ its moral properties and these moral properties consist of the non-moral properties.[47] On this version of appraiser internalism, motivation is not tied merely to the judgment that some Φ has a moral quality, but to the judgment that Φ has a moral quality because it has some non-moral quality. These are what she describes as the conditions under which someone comes to judge that an action or character trait is morally justified. The appraiser internalist thesis that Radcliffe attributes to Hume, then, is that "the conditions under which one judges that an action or a character trait *is or is not justified* are identical to the conditions under which one has a motive towards or away from that action or trait."[48]

---

[45] Radcliffe 1996, 389. I will, for reasons I explain presently, treat this as equivalent to weak motive appraiser internalism as that view is defined in the introduction.

[46] Ibid., 393; emphasis added.

[47] Ibid., 389. This claim is admittedly less than clear. It may amount to nothing other than the claim that the moral properties reduce to certain non-moral properties.

[48] Ibid.; emphasis added.

Now, Hume does not think that the conditions under which one judges that a character trait or action is morally justified are identical to the conditions under which a character trait or action are actually morally justified. The moral sentiments are particular kinds of passions, with their own distinct feelings, that arise under contemplative conditions of impartiality. Hume's sentiment-based account of morality does not make moral matters depend on an individual's experiences of just any pleasures or pains: "Nor is every sentiment of pleasure or pain, which arises from characters and actions, of that *peculiar* kind, which makes us praise or condemn…'Tis only when a character is consider'd in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil" (T 3.1.2.4, SBN 472). When I consider the threat posed to my happiness by my enemy I may initially feel a strong disapproval of him, and may on that basis come to think he has a "real villainy or baseness" (ibid.). However, since this displeasure arises from my regard to my own interest alone, it indicates that I am not "a man of temper and judgment" and so my "sentiments, from interests and morals, are apt to be confounded, and naturally run into one another" (ibid.). I may believe that character traits and actions are morally justified when they give me pleasure or take away my pain, and given this belief I have a motive to act from my self-interest. However, I do not in this case form any moral evaluations, since, for Hume, moral sentiments do not arise from such partiality.

Hume does not think that removing the conditions of partiality, where I am merely concerned with how various traits and actions affect my own welfare, is sufficient for moral evaluation. I may more strongly disapprove of my neighbor's suicide than I do of the suicide of a complete stranger, even though their deaths in neither case affect me at all. I may have a much stronger approval of my country's ever so slight increase in voter turnout than I do towards a neighboring country's establishing democratic institutions. I may feel more pleasure from the successes of my grandfather than I do towards anyone on Time's list of the 100 Most Influential People of the 20th Century. Hume would understand all of these cases as arising from the variable responses of the principle of sympathy. Hume describes sympathy as a natural

mechanism by which we can "receive by communication [other people's] inclinations and sentiments, however different from, or even contrary to our own" (T 2.1.11.2, SBN 316). When I observe certain behaviors in others, I infer that their behaviors are caused by particular feelings, perceptions, and dispositions. I make this inference on the basis of the fact that others are not so different from myself: we all have roughly the same psychological constitution, at least when it comes to the general principles by which our perceptions and behaviors operate. In addition to allowing me to form ideas of the inner-lives of others, our rough similarities also calls to my mind the lively perception that I always have of myself.[49] The idea of the other person's inner-life is enlivened by the perception I have of myself and "is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself" (T 2.1.11.3, SBN 317). Hume thinks that this transfer of vivacity from my perception of myself to the idea I form of other's passions is a matter of degree. The resulting feelings I have after this transfer depend on the relationship between myself and the person whose passions I am contemplating. The closer we are associated in terms of contiguity, resemblance, and cause/effect (familial relations), the more strongly I will be affected by the other person's passions. The vivacity (violence) with which I feel another's passions varies according to our nearness to one another (both in terms of time and space), our resemblance (extending to our appearances, languages, cultural practices, etc.), and our familial ties.

Hume recognizes that, insofar as he thinks our moral sentiments arise from sympathy, he is open to the following objection: "that these sentiments, whence-ever they are deriv'd, must vary according to the distance or contiguity of the objects; nor can I feel the same lively pleasure from the virtues of a person, who liv'd in *Greece* two thousand years ago, that I feel from the

---

[49] The lively perception that I have of myself is an important component in understanding how the idea I have of another's inner-life is converted into an impression. Hume says that "the idea, or rather impression of ourselves is always intimately present with us, and…gives us so lively a conception of our own person [that] whatever object, therefore, is related to ourselves must be conceiv'd with a like vivacity of conception according to the foregoing principles" (T 2.1.11.4, SBN 317). See Cohon 2008, 129 for a helpful discussion of the operation of sympathy.

virtues of a familiar friend and acquaintance. Yet I do not say, that I esteem the one more than the other" (T 3.3.1.15, SBN 581). Our moral evaluations of character traits remain stable in a way that the shifting intensity of the passions that arise from the operation of sympathy does not seem to allow. Hume continues:

> Besides, every particular man has a peculiar position with regard to others; and 'tis impossible we cou'd ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual *contradictions*, and arrive at a more *stable* judgment of things, we fix on some *steady* and *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation. (T 3.3.1.15, SBN 581-2)

The common point of view is not a vantage point where we can occupy some general and fully detached perspective.[50] Rather, we take up the particular perspectives of those individuals that will be directly affected by the person whose traits we are evaluating, including the person him or herself. We imaginatively project ourselves into these particular perspectives and use causal reasoning to discover how the traits in question would *likely* affect people so situated with respect to their pleasurable and painful states. When I infer that a trait or action would likely cause pleasure, I sympathize with the beneficiaries of that pleasure. When I infer that a trait or action would likely cause pain, I sympathize with the sufferer of that pain. These feelings of pleasure and pain in me just are my feelings of moral approbation and disapprobation.

This account of the common point of view presents a number of problems for Hume. Recall that Hume claims that "morality…is more properly felt than judg'd of" and that "in feeling that [some trait] pleases after such a particular manner, we in effect feel that it is virtuous" (T 3.1.2.1, SBN 470; T 3.1.2.3, SBN 471). With the account of sympathy generated from the common point of view, Hume is saying that the relevant pleasures for moral evaluation are the ones that arise from this perspective. Yet, it is unclear that these feelings, or any feelings

---

[50] Cohon says, "the common point of view is not a detached perspective, but the vantage point of the person being evaluated and the particular individuals with whom he has direct dealings. It gives us not a wide panorama, but an intimate glimpse" (2008, 144).

at all for that matter, arise from the general point of view. There are a number of passages which suggest that what determines our moral evaluations are not our *actual* sentiments, but are instead sentiments that we *would have* were we to actually take up the common point of view. Hume says that taking up the common point of view does not always affect our actual passions:

> But however the general principal of our blame or praise may be corrected by those other principles, 'tis certain, they are not altogether efficacious, nor do our passions often correspond entirely to the present theory. 'Tis seldom men heartily love what lies at a distance from them, and what no way redounds to their particular benefit; as 'tis no less rare to meet with persons, who can pardon another any opposition he makes to their interest, however justifiable that opposition may be by the general rules of morality. (T 3.3.1.18, SBN 583)

Hume adds that the "passions do not always follow our corrections; but these corrections serve sufficiently to regulate our abstract notions, and are alone regarded, when we pronounce in general concerning degrees of vice and virtue" (T 3.3.1.20, SBN 585). These passages, among others,[51] suggest that Hume does not think that our actual present feelings are what determine moral evaluations: the actual feelings that result when we take up the common point of view are not always the ones that we predict would arise. What these passages suggest, then, is that we "correct" our present actual feelings in light of our ability to infer how we would feel if we were situated in the common point of view, and the fact that our actual passions remain relatively intractable is irrelevant for the purposes of moral evaluation.

The problem with the internalist interpretation for which I have argued is now readily apparent. Consider again the case of the negligent father as an illustration of the Sense of Duty Passage. The negligent father lacks the common motive of natural affection for his child. He is in full possession of his moral sense and his awareness that he lacks this character trait gives rise in him to the feeling of disapprobation, though it is the feeling of self-disapprobation that we've been calling self-hate. I argued that this feeling is itself motivational. But given Hume's account of the operation of sympathy in the generation of moral evaluation, one might wonder if this

---

[51] T 3.3.1.16, SBN 582; T 3.3.3.2, SBN 603; EPM 5.41, SBN 227; and EPM 5.42, SBN 229.

description adequately fits the case. Since the negligent father is making a moral evaluation (he is making a judgment about his moral obligations), we suppose that he is occupying the common point of view. In theory, given the way in which sympathy is supposed to work from the common point of view, the father sympathizes with the pain that his negligence causes in his child and thereby comes to share in his child's sentiments. But might this not be a case where his actual passions fail to follow these "corrections"? If Hume holds the view that what is required for moral evaluation is merely that one take up the common point of view and apprehend hypothetical motives, then there is no reason why the negligent father couldn't both grasp his moral obligation to provide care yet fail to feel anything besides his own selfish motives. And if this is the case, then it is possible for him to understand his moral obligation but lack a motive to act accordingly.

On this interpretation, it seems that moral evaluation would not be determined by actual sentiments, but would instead be determined by judgments, based on causal inferences, about how one would feel if she were situated differently than she is.[52] I leave this worry aside here. My concern is what this interpretation suggests about the relationship between one's understanding of one's moral obligations and the motivations that she has to keep them. Suppose we change the negligent father case in such a way that he actually fails to act according to his duty. Suppose, further, that his failure to act is not the result of his being externally encumbered in some way.[53] Instead, he occupies the common point of view, judges correctly that if he were in his child's shoes he would experience pain in these conditions of neglect. However, because, we're supposing, this is a case where his own passions are not engaged by this consideration, he fails to be moved to act according to his moral judgment. This is precisely the possibility that an internalist interpretation is meant to exclude. If adopting the common point of view allows for

---

[52] Cohon 2008, 127.

[53] If this were the case, then Hume would say that we "esteem him virtuous, and are delighted with the view of his character, even tho' particular accidents prevent its operation, and incapacitate him from being serviceable to his friends and country" (T 3.3.1.19, SBN 584).

moral evaluation to be divorced from actual sentiments in this way, then it seems we have less reason to accept an internalist account than we did before.

Internalism and the Common Point of View: Possible Solutions

If Hume does accept a version of internalism, and his theory is not inconsistent, then what we need to see is how Hume understands the relationship between motivating sentiments and the moral evaluations that arise from the common point of view. I will here present two different ways of understanding how these views can be rendered consistent. Since both interpretations are consistent with internalism, I will not here argue in favor of one interpretation over the other. My aim in this section is just to show that an internalist interpretation can be reconciled with all that Hume has to say about the common point of view.

I will call the first solution the *approximating-sentiment view*.[54] This view takes as its starting point a claim that Hume makes in the *Enquiry Concerning Human Understanding*. Hume says, "a man of mild manners can form no idea of inveterate revenge or cruelty; nor can a selfish heart easily conceive the heights of friendship and generosity" (EHU 2.6, SBN 20).[55] He suggests that a person cannot really understand what it is for someone to have certain moral virtues or vices without also having sentiments that roughly correspond to the relevant virtues and vices.[56] A truly kind person cannot really understand what it is for someone to be cruel, and a thoroughly selfish person cannot really understand what it is for someone to be generous.[57]

---

[54] This is the name I am giving to Radcliffe's proposal. See Radcliffe 1994, 51-3.

[55] Hume uses this as an illustration of his claim that there can be no idea in the mind that does not have its antecedent source in some sense or other. A blind person cannot have ideas of colors and a deaf person cannot have ideas of sounds. The case presented here is meant to be analogous to these cases, and Hume says the sort of phenomena discussed here is present but in a lesser degree.

[56] As Radcliffe notes, this is distinct from the claim that someone cannot learn how people use moral terms without having certain sentiments. This seems to be possible on Hume's view, but the passage from the *Enquiry* suggests that *understanding* the moral terms cannot occur without certain sentiments.

[57] As an analogous case Radcliffe adds that a person cannot really understand a sophisticated musical score if she has never before heard a single note of music. See Radcliffe 1994, 51.

However, all that is required in order for someone to be in a position to be able to understand what it is to be a cruel person is that they have *some degree* of cruelty, and all that is required for someone to be in a position to be able to understand generosity is that they be to *some degree* generous.

Radcliffe then wants to understand the motivational states that arise from moral evaluation in an analogous way: one only need to have motives that *approach* or *approximate* the motives one would have from the common point of view. When I take up the common point of view and evaluate a cruel act committed by my dear friend, I may not feel the same intensity of disapproval towards her that corresponds to the vice of malice, and so my actual feelings are out of sync with the feelings that, say, a distant stranger would have towards her were the stranger to occupy the common point of view. Nevertheless, on the approximating-sentiment view, I still have some degree of disapproval towards my friend's act of cruelty, and this disapproval provides me with a motive to act. I do not actually feel the *proper* feeling, but I do have a feeling that approximates it and this feeling is sufficient both to provide the correct moral evaluation as well as sufficient motivation. We extrapolate our moral judgments from the feelings that we actually have, so that our moral judgments do not directly track our moral sentiments. This still counts as a sentimentalist interpretation of Hume, however, since a person "arrives at proper moral judgments while being affectively engaged by sentiments that approximate to those from which the moral judgments are projected."[58]

The approximating-sentiment view seems to hold that we feel no new sentiments when we engage the common point of view and are unable to shake our own intense partial feelings. This is not the case on the other possible solution I wish to discuss, the *double-sentiment view*.[59] Cohon notes that all of the passages where Hume says that we may be unsuccessful in bringing our actual feelings in line with the common point of view are consistent with the view that we

---

[58] Radcliffe 1994, 52.

[59] This is the name that Cohon 2008 gives to her interpretation.

feel some new sentiment in these cases. She describes the process of taking up the common point of view as follows: "we feel certain passions from our particular vantage point, and whenever we contemplate the same character from the common point of view, we feel *another*, weaker sentiment. That is, we feel two sentiments toward the same character trait."[60] Although Cohon here calls the feeling that arises from the common point of view a "weaker sentiment," she makes it clear that the feeling that is given rise to is a calm passion.[61] While we are able to use the calm passion, which arises from our occupying the common point of view, to convert our intense and partial feelings into the moral feelings themselves, this is not always the case. Where we cannot do so, we remain in an emotional state of conflict with respect to the trait in question: we have both a violent and biased feeling as well as a calm and general one towards the same character in question. It is only the calm and general passion that counts as the moral sentiment. Moral judgments are made on the basis of this sentiment alone, not on the basis of the more intense feeling arising from our particular situation. This feeling is present whenever we occupy the common point of view. Now, this feeling may be exceedingly weak in light of the other passions that we currently experience. However, all that is needed in order to reconcile the common point of view with internalism is that *some* motive or other is present in all cases of moral evaluation. This is clearly secured by the double-sentiment view.

The double-sentiment view is strengthened by some of Hume's own claims later in T 3.3.1. He says that, when our actual feelings fail to merge with the passions that arise from the common perspective, we think of these as cases where "*reason* requires an impartial conduct" even though our "passions do not readily follow the determination of our judgment" (T 3.3.1.18, SBN 583). By "reason," however, Hume means to recall his discussion from T 2.3.3 where the conflict between "reason" and passion was attributed to a conflict between calm and violent

---

60 Cohon 2008, 139-40.

61 That the sentiment is weak is a matter of its limited motivational influence, it is not an indication of its felt vivacity, which is the point at issue.

passions. By recalling this discussion explicitly, Hume suggests that these too are cases where there is conflict between a calm and violent passion, and this lends support to the double-sentiment view.

In either case, internalist interpretations can still succeed. Both the approximating-sentiment and the double-sentiment views show that when one makes a moral judgment, some affective state or other is involved. So, the difficulty that we imagined to arise from the negligent father case simply doesn't arise at all. If the negligent father understands his moral obligation, then he must feel something besides his own selfish motivations. He need not be *moved to act* in accordance with these feelings, but insofar as he in fact judges that he is morally obligated, he must have some sentiments in addition to his own selfish ones. It seems then that Hume is able to dissolve the apparent tension between internalism and the common point of view.

"Why Be Moral?"

There are now two main sets of questions that must be addressed when we consider Hume's response to the question "why be moral?" The first set of questions arises when we consider that the moral sentiments that are felt from the common point of view may not be particularly strong. Given internalism we have some reason or motive to do as we judge that we ought, but we might well wonder whether we ever have a sufficiently compelling reason to do as we judge that we ought. The next set of issues arises when we consider that projecting oneself into the common point of view is not something that happens without effort, and projecting oneself into this point of view is liable to give rise in one to various sorts of pains that she could easily avoid by simply never adopting the common point of view. One might wonder why she should bother adopting this common point of view in the first place. I turn to these two different sets of concerns in the following sections.

Motives for Acting Morally

On the interpretation of Hume that I have been defending, if someone judges that she is morally obligated to perform some action (because, say, she recognizes that failing to perform it

would cause impartial and sympathetic spectators to disapprove), then she would feel disapprobation towards herself upon the contemplation of non-performance. This self-disapprobation, either alone or in tandem with naturally virtuous motives, provides some motive for her to act as a virtuous person would. So, if Hume were to approach the question "why be moral", he would do so from within a framework that maintains that moral judgments themselves already provide some motive to act.

As we have seen in the previous section, it is clear that Hume does not think that the moral sentiments are always particularly strong motives. Additionally, Hume says, "*Thus self-interest is the original motive to the* establishment *of justice: but a* sympathy *with public interest is the source of the* moral approbation*, which attends that virtue*. This latter principle of sympathy is too weak to controul our passions; but has sufficient force to influence our taste, and give us the sentiments of approbation or blame" (T 3.2.2.24, SBN 499-500). As Radcliffe explains, this is why Hume spends time detailing various external incentives towards virtuous behaviors, including the pleasurable feelings instilled through political speeches, public praise and blame, private education, and the positive effects on our own reputation of our just behavior.[62] Nature may "furnish the materials, and give us some notion of moral distinctions" but it is the "artifice of politicians," among other things, that helps to produce "those sentiments, which [nature] suggests to us, and may even on some occasions, produce alone an approbation or esteem for any particular action" (T 3.2.2.25, SBN 500). This suggests that, because acting as morality requires is pleasant and pleasure is productive of desires, the pleasures of moral action produce supplementary and strengthening motives to act in ways we judge that we ought to.

So, from the perspective of *motive* internalism, Hume can only address the "why be moral" question insofar as that question concerns *strengthening* motives that one already has in virtue of making moral judgments. But now we need to ask whether or not acting as morality demands is *reasonable*, and to see what Hume would say when we ask if we have some

---

[62] See Radcliffe 1996, 404.

justifying reason for acting as morality requires. Recall that Hume has very strong commitments regarding the rationality of action. He, like Hutcheson, thinks that reason is nothing other than the discovery of truth and falsity. However, unlike Hutcheson, Hume does not draw a distinction between exciting and justifying reasons and so does not seem to be able to talk about any literal sense in which an action counts as being reasonable or unreasonable. In order for something to count as being reasonable on Hume's view, it must admit of the sort of *agreement* or *disagreement* relations constitutive of truth or falsity. Passions and actions, however, do not admit of any agreement or disagreement with relations of ideas or matters of fact. As I mentioned in the previous chapter, I think the basic thought here is that, because passions and actions are not *copies* of other mental phenomena, they can be neither accurate nor inaccurate copies of their originals. Thus, they cannot be either true or false, and so cannot be reasonable or unreasonable. There does not then seem to be a sense in which Hume can say that one has a justifying reason to be moral; one has no justifying reasons for acting at all.

However, Hume does mention two *derivative* senses in which an individual may be said to be acting unreasonably. The first is when she has false beliefs about the existence of an object of her desire, and the second is when she has false beliefs about the necessary means to acquire the objects of her desires or passions.[63] It is not at all clear that vicious and immoral actions will be unreasonable in either of these two derivate ways. Hume explicitly mentions a few examples of actions that are pretty clearly immoral and vicious (on his account) that he nevertheless thinks it is not unreasonable to perform. Here are the cases:

> 'Tis not contrary to reason to prefer the destruction of the world to the scratching of my finger. 'Tis not contrary to reason for me to choose my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. (T 2.3.3.6, SBN 415-6) [64]

---

[63] Radcliffe 2013, 17-20 provides a helpful analysis of these two derivative senses in which an action or a passion can be unreasonable.

[64] Weller 2004 provides a very close and detailed account of these cases.

These cases are meant to be surprising: Hume is drawing out the rather controversial implications of his theory of practical rationality.

Each of these cases involves an agent more strongly desiring some good that she recognizes is the lesser good compared to another option. The first case involves someone having a stronger desire to avoid a trivial injury to her finger than for the good of the continued existence of the world. In the second case, a lesser good than one's own life is desired more strongly (as evidenced by its being the option that one chooses) than a rather trivial good to a complete stranger. Finally, in the third case, the stronger desire attaches to one's own lesser good that one recognizes is one's own lesser good. What seems to be irrational, in our ordinary way of speaking, about these cases is that they are cases where one's recognition of the value of things is out of sync with the strength of one's desires for them. However, for Hume, these desires are not irrational since they in no way involve a false belief. He goes on to provide an account of how the lesser good can give rise to stronger motives than recognized goods, and this account appeals merely to the intensity of present pleasures and pains compared to distant pleasures and pains. The present pleasant feeling I get from sending all of my money to a stranger is stronger, and the idea of it more vivid, than any idea I have of the pleasures that would be available to me in the future if I kept the money. It is also strong enough to outweigh the rather faint idea of pain that I have presently contemplating the suffering that will result from my having no money.[65] What we would ordinarily describe as irrationality is, for Hume, simply a matter of the relative force and liveliness of our ideas of the pleasures and pains that do or would arise from different courses of action. If there are no false beliefs involved in these preferences, then they are not contrary to reason (and if there were false beliefs involved, it would be these beliefs, not the preferences, that would be contrary to reason).

However, as Radcliffe argues, the mere fact that his preferences and his choices are not irrational (because they are neither rational nor irrational) does not mean that there are no other

---

[65] Radcliffe 2013, 22.

grounds on which an agent can be evaluated. In fact, as I mentioned above, it seems fairly clear that these are all cases of vicious actions. The first case involves someone preferring utter destruction to the entire world just for the sake of avoiding a very trivial harm to himself. This preference would not be approved from a *disinterested* perspective, since the strength of the desire seems to depend essentially on the agent's interest in avoiding the cut on his finger.[66] Similarly, a disinterested spectator would not approve of sacrificing one's own greater good for the lesser good of another, especially not for the sake of a stranger. The problem here is that developing a disposition to sacrifice oneself for rather trivial gains is not useful, at least, not as useful as developing a disposition to only sacrifice oneself when the gains are sufficiently high. By taking up the common point of view we would feel some sympathetic pleasure upon contemplating the improvement made to the stranger's condition. However, we would also feel a much stronger sympathetic pain with our own losses and with the effects these losses would have on those nearest to us. The final case, the case of imprudently preferring one's lesser good to one's acknowledged greater good, might not seem like a moral matter. However, Hume does treat prudence as a virtue. He says,

> [i]f we examine the panegyrics that are commonly made of great men, we shall find, that most of the qualities, which are attributed to them, may be divided into two kinds, viz. such as make them perform their part in society; and such as render them serviceable to themselves, and enable them to promote their own interest. Their prudence, temperance, frugality, industry, assiduity, enterprize [*sic*], dexterity, are celebrated, as well as their generosity and humanity. (T 3.3.1.24, SBN 24)

He adds that imprudence is an "imperfection of character" that is so serious that "men wou'd rather acknowledge the greatest crimes, than have it suspected, that they are, in any degree, subject to [it]" (ibid.). So even if it is not unreasonable for one to prefer her own present lesser

---

66 As Radcliffe and Baier have noted, the phrase "the destruction of the world" seems to indicate that this finger-scratch-averse person would also be destroyed. So, perhaps as Baier suggests, he has a spaceship to carry him out of the world before its destruction. Alternatively, if we don't want the case to build in any false beliefs, he may simply prefer to be annihilated himself, along with everyone else, rather than to endure the misery that accompanies his paper-cut, or whatever.

good to her acknowledged greater future good, for her desires to be arranged in this sort of a way is a mark of moral imperfection in her.

Hume can, and does, say that being virtuous is part of one's happiness. As we have seen, he says that an "[i]nward peace of mind, consciousness of integrity, a satisfactory review of our own conduct – these are circumstances very requisite to our happiness" (EPM 9.2.23, SBN 283). So, the agent acting from the not unreasonable, but likely vicious, desires mentioned above may acquire some motive for acting as morality requires. She may also acquire a motive to rearrange her desires. The fault for which she can be blamed is a moral fault. It is not a rational fault. She has no reason, outside of better satisfying her preexisting desire for happiness, to strengthen her desires to do what morality requires, or at least to avoid doing what morality forbids. Hume can talk about having motives to act morally, and can give an account of how these motives can be strengthened. However, Hume cannot provide an account of why it is reasonable for one to act in this way, or why it's unreasonable for one to fail to act in this way, unless we can somehow tie these cases to false beliefs.

"Why Should I Take Up the Common Point of View?"

The question that arises now for Hume is a question that we saw also arises for Hutcheson. Hutcheson recognizes that benevolence is a multifarious phenomenon, and that our particular benevolent affections tend to be stronger than our universal benevolence. Moreover, since our will is determined by those affections that are the strongest, it seems that our natural motives will not always guide us towards the benevolent actions that are partly constitutive of the most virtuous actions available to us, namely those motivated by universal benevolence. However, by undergoing a process of "enlarging and abstracting," there arises within human agents the calm universal benevolent motives, and acting on these motives does direct us towards those acts that are the most virtuous. So, for Hutcheson, the question becomes why we should engage in this process of "enlarging and abstracting." I suggested above that Hutcheson's view is that we have reasons from self-love to engage in this process, even though self-love cannot

originally cause a motive to virtue. Since benevolence is our own greatest happiness, and cultivating an extensive benevolence is the only way to consistently receive the pleasures that arise from a reflective self-approbation, self-love serves as a reason to *strengthen* benevolence, or, rather to widen its scope.

Even though it will be different in its details, I think that Hume will give the same sort of response if we ask him why we should bother taking up the common view at all. Given Hume's internalism, when I make a moral judgment, I also have certain motivating passions. We have also seen that acting from these motives helps me to secure certain desires that I have, and so I have some motivation to try to strengthen the motives that arise from my having made a moral judgment. But my making a moral judgment is something that requires deliberate effort on Hume's view. The common point of view isn't one I just so happen to occupy. It takes quite a bit of imaginative work, and also requires that I carefully employ various empirical beliefs that I have in order for me to adopt it. What is more, when I occupy the common point of view I open myself up to experience pains that I would otherwise be able to avoid. Taking up the common point of view does not merely give me an *idea* of the pains that others feel when they interact with my own or other's vicious characters. I feel pain via the operations of sympathy when I imaginatively put myself in the shoes of those in closest contact with vicious characters.

Let's return to consider another variation on the negligent father case. As I developed that case with respect to the Sense of Duty Passage, the negligent father comes to act from a sense of his duty after apprehending his moral obligation. This means that the father has taken up the common point of view and felt sympathetic pains strong enough to move him to act. It is true that his acting from the sense of duty may give him pleasure, just as the person who lacks strong generous motives feels pleasure on acting as the generous person would (T 3.2.1.8, SBN 479). Suppose that the pleasure that he feels is not particularly intense and that he nevertheless remains acutely aware of his lacking the naturally virtuous motives since, we'll suppose, he is unable to disguise from himself his lack of natural affection. Hume clearly thinks that in this case the consistent self-disapprobation will continue to goad the father to fulfilling his parental duties

until he acquires the natural virtue through habit. However, it still seems open for the negligent father to respond differently. He may instead come to realize that the pains that he feels arise only when he imagines himself in his child's shoes and makes causal inferences about how he would feel from that point of view. He may instead simply refuse to adopt the common point of view. This explains why Hume says that "a person, who feels his heart devoid of that principle [common in human nature], *may* hate himself upon that account" (T 3.2.1.8, SBN 479; emphasis added). He will only hate himself (i.e. feel moral self-disapprobation) if he projects himself into the common point of view. He may even attempt to justify his refusal by appealing to his own past experiences or to the child's own good. He may point out that no one ever took his own interests to heart and he turned out just fine, or that occupying such a perspective is coddling the child, which would stifle the development of independence. What, if anything, can Hume say to such a character to convince him that he should take up the common point of view?

There is one sort of a response that was available to Hutcheson that is not available to Hume. Hutcheson thinks that the facts of our natural benevolence ground the reasons why we should act benevolently. We saw above that we have reason to be virtuous because doing so satisfies interests that we already naturally have. This response will not work with respect to the negligent father case. The point that is at issue arises precisely because the father lacks the naturally motivating sentiments. Acting as the caring parent would does not satisfy his already present interests because he does not have natural affection in the first place. His reason for taking up the common point of view cannot be because doing so allows him to realize something that he already has an interest in realizing, namely the well-being of his child for its own sake.

We supposed that the negligent father may attempt to justify his refusal to sympathetically engage with his child's point of view by pointing to his own experiences or to some good for the child. Suppose that he were even to try to frame this attempted justification in moral terms. He might say, "My parents did right by me and I am simply doing as they did," or again "It is wrong to stifle a child's independence by coddling them." It is open to Hume to

charge the negligent father with a failure to correctly use moral language. Hume says in the second *Enquiry*:

> When a man denominates another his *enemy*, his *rival*, his *antagonist*, his *adversary*, he is understood to speak the language of self-love, and to express sentiments, peculiar to himself, and arising from his particular circumstances and situation. But when he bestows on any man the epithets of *vicious* or *odious* or *depraved*, he then speaks another language, and expresses sentiments in which he expects all his audience are to concur with him. (EPM 9.6, SBN 272)

By refusing to take up the common point of view, the negligent father is not actually able to use moral language. Since moral language expresses sentiments that we expect our audience to share, we can only correctly employ such language when we take up the common point of view. Otherwise, our judgments will be based on our own partial feelings rather than on those of others.

However, it is open to the negligent father to opt out of using moral language altogether. This "enlightened" negligent father will realize that he is not really entitled to make moral judgments without adopting the common point of view. So, he can evade the charge of misusing moral language by not using it at all, and continuing to refuse to adopt the common point of view. The enlightened negligent father will have no purported moral justification for his refusal. The only thing he will be able to point to is the fact that taking up the common point of view will cause him pain, and he desires to avoid pain.

He can still refuse to use moral language and to adopt the common point of view even if it is true that, in general, adopting the common point of view has practical advantages. For instance, there are regular causal connections between one's judgments about a person's virtues and vices on the one hand, and the likelihood that association with that person would give rise to pride or humility, or love or hatred.[67] Hume says that "these two particulars are to be consider'd equivalent, with regard to our mental qualities, *virtue* and the power of producing love or pride, *vice* and the power of producing humility or hatred" (T 3.3.1.3, SBN 574-5). Recall that the

---

[67] Cohon 2008, 144-58.

common point of view was developed in order to account for the stability of our moral judgments despite the fact that they arise from the operation of sympathy which shifts its deliverances according to contiguity, resemblance, and cause and effect. The closer, in any of these senses, we are to someone the stronger they will engage our sympathies. I will be much more strongly affected by my city council's poor decision to invest in a hideous public work of art than I will be by a recent homicide committed in Siberia, but judge the homicide to be morally much worse. Hume explains this by saying that I morally judge (and so feel, in one of the ways I specified above) from the common point of view. If I were to judge from my own particular station and perspective, it seems I would be committed to maintaining that the frivolous spending of my city council is a more egregious moral failing than the homicide. From my own perspective, I may feel an intense hatred towards my city council and humility towards myself for being at all associated with them. My feelings of hatred towards the homicidal Siberian, however, won't be as intense and I would likely feel no humility at all since any relation he has to me is rather tenuous. I might come to prefer the company of the Siberian murderer to the company of any of my city council members, and come to desire taking up residence with him over continuing to reside in my own city.[68] However, it seems fairly clear that were I to judge characters from my own perspective alone, ignoring the perspective of the murderer's victim and his family, I would end up making grave mistakes about the people with whom I choose to associate. In order for me to make good predications about those associations of mine that will likely give rise in me to pride and love, then, I need to take up the common point of view and see how the traits of the person in question are likely affecting the pleasures and pains of those with whom he has the closest interactions.[69]

The common point of view, then, is generally useful for me to employ as a tool to arrive at true beliefs, namely true beliefs about the traits that count as virtues and vices, and true beliefs

---

[68] For similar sorts of cases see Cohon 2008, 147.

[69] Ibid.

about the likelihood that certain people will cause me either pride or humility, or for whom I will develop either love or hatred.[70] These beliefs are important for me to have because they help me be judicious in selecting my associates so that I can avoid those people who will likely cause me pain and pursue those people that will likely cause me pleasure.[71] Adopting the common point of view, then, is often a matter of convenience for us. It is because the common point of view is convenient and useful for us in forming beliefs that Hume is able to agree with part of Hutcheson's account: we usually have some reasons of self-love to expand our circle of concern beyond what we're naturally and immediately affected by (or affectionate towards in Hutcheson's case). The enlightened negligent father, who no longer frames his refusal to adopt the common point of view in moral terms, can recognize this convenience and can even adopt the common point of view in a great number of other cases. However, the pain that he experiences when he adopts his child's perspective might just be too intense and the effort required to actually establish the virtuous motive too great, that he simply refuses in this case to think about his child's well-being. Now, a child raised by such a person would surely grow up resenting him and so it's possible the child will make sure the father experiences pain once he's grown. If this were the case, then the father would merely be deferring his experiences of pain. He may even be aware of this and so would be plausibly described as preferring his short term good to his acknowledged long term good. This, of course, is not unreasonable on Hume's view. So long as there are no false beliefs involved, there is nothing unreasonable about his actions. Now, suppose it's true that neglected children are likely to seek revenge on their parents, and suppose the father comes to learn this. He might be supplied with a motive strong enough to overcome his resistance to adopting the common point of view. However, he might not and there doesn't seem to be anything we can say to him to show him that he *should* take up the common point of view.

---

[70] "Virtue is consider'd as means to an end. Means to an end are only valu'd so far as the end is valu'd." (T 3.3.6.2, SBN 618-9).

[71] This important point is developed by Cohon 2008, chapter 4.

Note that the self-interested story as I have developed it in this case crucially depends on the father's belief of the likelihood that the child will end up causing him pain. This detail is not essential to my overall point. We may, for instance, try to convince the father of the point Cephalus makes at the beginning of *The Republic*, that old age is more burdensome for the wicked person than it is for the good person. We could convince him that the elderly tend to have fewer commitments, more time to reflect, and fewer opportunities to pursue new experiences. All of this leads naturally to thoughts about what might have been. Given the strength of the parent/child relationship, even if the strength of this relationship is not now apparent, it might manifest itself later in these reflective conditions. We might convince him that most negligent parents feel a great deal of remorse and regret that they weren't better parents. Even if we are successful in convincing him of this, we cannot call his refusal to take up the common point of view in light of it unreasonable. This new belief might move him to begin taking his child's interest into consideration. But it might not and when it doesn't there is nothing, it seems, we can say to convince him that he should.

A Moral Obligation to Adopt the Common Point of View?

Perhaps we could appeal to the moral sentiments themselves in order to explain why it is that he should bother with thinking morally about his child. The idea here is that we could appeal to Hume's account of moral obligation and argue that the sense in which one should adopt the common point of view is that one is morally obligated to do so. To say that someone is morally obligated to adopt the common point of view is just to say that impartial and sympathetic observers would disapprove *from the common point of view* that person failing to adopt the common point of view. Since Hume's view of obligation is sentimentalist in this way, it is not entirely clear that he can say there is a moral obligation to take up the common point of view.

There may be cases where we approve of one's motives even when the agent did not take up the common point of view. The person who naturally possesses "meekness, beneficence, charity, generosity, clemency, moderation, equity" and who is disposed to act according to the

conventions required for the artificial virtues may never actually adopt a common point of view when she deliberates about what to do (T 3.1.1.11, SBN 578). Hume seems committed to the view that we would approve of the agent that is in possession of these virtues regardless of whether her deliberative procedures appealed to the common point of view or not. Hume insists that all of the virtues depend on sympathy: "we shall not doubt, that sympathy is the chief source of moral distinctions…[all virtues] must derive all their merit from our sympathy with those, who reap any advantage from them: As the virtues, which have a tendency to the good of the person possess'd of them, derive their merit from our sympathy with him" (T 3.3.6.1, SBN 618). Our sympathy is engaged when we consider *the effects* that certain traits have on others (or on their possessor), and we denominate as *virtues* those traits that are useful or agreeable to society (or their possessor). Impartial and sympathetic observers approve of the person that is naturally sound, that is, in full possession of the natural tendencies to produce social and personal goods and this is so even if she herself were never to make moral judgments or take up the common point of view at all.[72] Even if she were to adopt the common point of view, it seems fairly clear that she would not disapprove of her natural motivations, aligning as they do, with the good of herself and others. So, it is not clear that Hume can maintain a general moral obligation to take up the common point of view.

Perhaps, then, we could argue that the negligent father is failing in some more particular moral obligation. Let's grant that he is not under a general moral obligation to take up the common point of view. He might still have a moral obligation to not avoid taking up the common point of view as a means to avoid feeling pain. This is not a general obligation to take up the common point of view under any and all circumstances. The obligation in this case is an obligation regarding the means by which one should avoid feeling pain: it is morally wrong to avoid feeling sympathetic pains by refusing to put oneself in the conditions under which sympathy can operate. I think that it's plausible that we could consistently maintain that there is

[72] For a similar point see Brown 1988, 81.

such an obligation on Hume's view. As impartial and sympathetic observers, we are likely to strongly disapprove of the father's stopping his ears at the suffering of his own child. Moreover, given that the case in question involves the father having discovered that occupying the common point of view is a source of uneasiness for him, it seems clear that the father too would disapprove of his omissions were he to dare to project himself into this perspective. But the father's response to this moral obligation is all too obvious. He will simply tell us that we have given him even more reason to refuse to take up the common point of view. In addition to the sympathetic pains that he will feel when he projects himself into his child's position, he will additionally feel pains arising from his own prior refusal to having taken up that perspective. He will only agree with us that he is morally obligated to avoid refusing to take up the common point of view if he first takes up the common point of view, and he wants to know why he should bother. Our telling him that if he does so he will feel uneasiness seems unlikely to be compelling.

I do not think that we can avoid concluding that, for Hume, there is not much that can be said in order to convince the negligent father to engage in the practice of moral reflection. It is clear that, given Hume's internalism, *if* the father were to engage in that practice, he would have some motive to act contrary to his own natural inclinations. I think Hume can also say that *if* he takes up the common point of view, he will be strongly motivated to continue to develop virtuous dispositions since doing so will ultimately give him more pleasure than pain. Yet, when we are dealing with someone who refuses to take up the common point of view, even if this is someone who is only refusing to take up this perspective in certain isolated cases, so long as no false beliefs are involved, there is nothing we can say to convince him that he is being unreasonable.

Concluding Remarks

I think it would be incorrect to see these reflections on the negligent father as showing a limitation of Hume's theory. The negligent father is not a sociopath, because he is capable of sympathetic pains and pleasures. Yet, I suspect that the sort of figure that we are imagining is not

one that Hume thinks is very pervasive in human society. Hume seems to think that most people, even if they are unable to consistently act virtuously, nevertheless make moral judgments and even approve of their doing so. He says that, "all lovers of virtue (and *such we all are in speculation*, however we may degenerate in practice) must certainly be pleas'd to see moral distinctions deriv'd from so noble a source" and he add that "a sense of morals is a principle inherent in the soul, and one of the most powerful that enters into the composition" (T 3.3.6.3, SBN 619; emphasis added). Given the kinds of creatures that we are, asking why we should *think* morally is typically moot.

For Hume, the ultimate ground of our normative judgments is the approval of our moral sense. Our moral sense determines what is virtuous and what is vicious. But it also approves of itself, as well as the principles by which it operates: "According to the latter system [Hume's], not only virtue must be approv'd of, but also the sense of virtue: And not only that sense, but also the principles, from whence it is deriv'd" (T 3.3.6.3, SBN 619).[73] There is nothing beyond the approval and disapproval of the moral sense to which one could appeal to say that we should or should not act in certain ways. When we thus choose to suspend the operation of our moral sense, by say refusing to imaginatively project ourselves into the common point of view, the resources of Hume's theory simply run out.

I suspect that Hume would welcome this result. There is only so far that philosophy is able to go. When Hume rhetorically asks, just before introducing the famous painter/anatomist analogy, whether there be anyone who doesn't feel an "accession of alacrity" when he considers that his natural virtues bring him "a new lustre in the eyes of mankind," he knows that there are some who do not (T 3.3.6.6, SBN 620). Similarly, when he asks if anyone really thinks that acting against a social virtue is ever worthwhile given that "a mind will never be able to bear its own survey, that has been wanting in its part to mankind and society," he knows that there

---

[73] For more on this reflexive approval as a source of normativity for Hume, see Baier 1999, chapter 9; Korsgaard 1996; and Cohon 2008, chapter 9.

certainly are such individuals (ibid.). But in response to such possibilities he says, "I forbear insisting on this subject. Such reflections require a work apart, very different from the genius of the present" (ibid.). Hume seems well aware that he only has so much to say to those who insist on refusing take up the common point of view. The work that he has done in moral philosophy can only go so far towards persuading people to act in certain ways.

I do not want to offer a general interpretation of the painter/anatomist analogy, nor do I wish to say that Hume was taking himself to be acting in the role of one or the other, or both. Instead, I want to suggest that when Hume appeals to the role of *painter*, he is giving us some clues about "the work apart" to which he previously appealed. To respond to those who refuse to engage in moral practices, Hume seems to be suggesting that what they need isn't a moral theory. They do not need to be *convinced* of anything. Rather, what they need is to be *influenced*. They need to see that moral practices are beautiful, awesome, inspiring, powerful, and compelling. What they need cannot be provided by moral philosophy, or by any theoretical approach at all. However, great works of art can supply what they need. There are no arguments to persuade the negligent father to begin to think morally. However, he will find resisting such thinking rather difficult when he's exposed to a poetic composition about child abuse, or faced with an artist rendering of a suffering child. While detailing how such effects would come about on Hume's psychology is not something I can do here, I am confident that Hume's theory has the resources to develop such an account. Works of art will play an important role in influencing and inspiring people to take up the common point of view. The painter/anatomist analogy thus leads us to believe that works of art will be an important part of what Hume calls practical morality.

These limitations to moral theory are well recognized by Hume, as evidenced by the concluding remarks of the conclusion of Book 3: "And thus the most abstract speculations concerning human nature, however cold and unentertaining, become subservient to *practical morality*; and may thus render this latter science more correct in its precepts, and more perswasive [*sic*] in its exhortations" (T 3.3.6.6, SBN 621).

CHAPTER 7

EMPIRICAL NATURALIST INTERNALISM?

Introduction

I have not as of yet said anything about Darwall's interpretation of Hume. This may seem a glaring oversight given the centrality that Darwall has in this dissertation. I have delayed a discussion of Darwall's more recent views of Hume because I think that they reveal something very important both about Hume and about Darwall's overall project. In the first section of this chapter I will present a critical interpretation of Darwall's reading of Hume. The main thrust of the criticism here is that Darwall, in framing his investigation of the British Moralists as he does, is searching for something that cannot be found in the empirical naturalists. And, what is more, this is no fault of these theories taken on their own terms. In the second section of this chapter, I will argue that Darwall's interpretation of Hutcheson, and to a lesser extent Hobbes, seems to read in too much of other figures views. Finally, I conclude with a discussion regarding the limitations faced by Darwall's taxonomy.

Darwall's Hume and Rule Obligation

In his discussion of Hutcheson, Darwall, as we have seen, is rather puzzled by Hutcheson's adoption of a sentimentalist account of obligation: "In disconnecting obligation from rational motive, Hutcheson implicitly rejected without any evident rationale the understanding of obligatory force that had become so widely accepted in early modern British ethics (especially among empiricist-minded philosophers)."[1] Because Hutcheson had available to him a view that clearly rooted obligation in rational motives, but rejected the view nonetheless, Darwall finds something odd about the theory that Hutcheson develops. Since Darwall thinks that Hume inherits Hutcheson's account of obligation, he thinks the same sort of puzzle arises with respect to Hume. However, Hume's account of justice departs from

[1] Darwall 1995, 234.

Hutcheson's, and Darwall seems to suggest that this account shows Hume recognizing limitations both with the sentimentalist account of obligation in particular, and with empirical naturalism in general.

In the next two sections, I will present two of the exegetical puzzles that Darwall uses to motivate his interpretation. Before doing so, I will present his argument for the claim that Hume develops a unique species of obligation in addition to recognizing the two kinds he inherits from Hutcheson. Darwall points to a passage where Hume says that obligation presupposes the establishment of convention: "Those, therefore, who make use of the words *property*, or *right*, or *obligation*, before they have explain'd the origin of justice, or even make use of them in that explication, are guilty of a very gross fallacy" (T 3.2.2.11, SBN 490-1). As we have seen, Hume takes on board Hutcheson's account of *natural obligation* as a motive of self-interest, as well as his account of *moral obligation* as a sentiment of moral approval (or disapproval of omission).[2] His doing so makes his use of the term 'obligation' in the T 3.2.2.11 passage rather odd. As Darwall points out, neither sense of obligation, natural or moral, depends on the establishment of social conventions. This leads Darwall to conclude that Hume must have in mind some third sense of obligation when he talks about the obligation to be just, to keep promises, and etc. Darwall calls this third species of obligation "rule obligation" and says that it is a sort of obligation that arises from the rules of justice themselves.[3] He uses this third kind of obligation to argue that Hume's view is less empirical naturalistic that is usually thought. In order to see this, we must return to examine Darwall's puzzles.

Hume's Theory of Will is Inconsistent with the Account of Justice

The first puzzle regards the relationship between Hume's "orthodox" theory of the will and his account of the artificial virtue of justice. We have already seen Hume's orthodox account

---

[2] See T 3.2.2.23, SBN 498 for a statement of Hume's commitment to these Hutchesonian senses of obligation.

[3] Darwall 1993, 425.

of the will. He says, for instance, "the will exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body," and he also says that "nature has implanted in the human mind a perception of good or evil, or in other words pain and pleasure, as the chief spring and moving principle of all its actions" (T 2.3.9.8, SBN 439; T 1.3.102, SBN 118). But the virtue of justice does not seem to be so narrowly focused on the production of good and the avoidance of evil. Darwall claims that this aspect of Hume's view is a departure from Hutcheson. Hutcheson maintains that the rules of justice are always to be subservient to the production of the common good. Hume however seeks to develop an account where people aim to "regulate their conduct by certain rules" (T 3.2.2.10, SBN 490). Darwall insists that the sorts of rules that people in society express an interest in following are not merely rules of thumb. They are constitutive rules.[4]

Darwall contends that an agent's regard for these constitutive rules is rather unique in Hume's overall psychology. Not only is it the case that the rules of justice are constitutive of a practice (of personal property, say), but each agent also takes the rules to be prescriptions for her own behavior. She regards them, as Darwall puts it, as being normative. Part of his argument for this claim is textual. He points to passages like the following:

> I observe, that it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually express'd, and is known to both, it produces a suitable resolution and behaviour. (T 3.2.2.10, 490)

---

[4] Darwall appeals to Rawls's distinction between summary and practice rules to make this point. To avoid any unnecessary complication of my own exposition, I have chosen to describe Darwall's view in terms that, I hope, both accurately portrays that account and carries a bit less interpretive baggage. By a 'rule of thumb' I mean, roughly, a rule the following of which tends to lead an agent to the successful achievement of a particular goal. "Do not use your Queen when a Bishop or Rook could do the same job" might be rule of thumb in chess. By a 'constitutive rule' I mean, roughly, a rule that defines a practice and that cannot be forgone without abandoning the practice itself. "Queens cannot move in the manner of Knights" is a constitutive rule of chess. Of course, one can move her Queen in an L-shape on a chessboard all she likes. But she cannot do so while continuing to play a game of chess. I do not intend for this to be an interpretation of Rawls (or Hart), but use it merely as way to present aspects of Darwall's view.

Darwall also cites places where Hume seems to say that agents adopt "general inflexible rules" that are regarded as "sacred and inviolable," and that everyone must "fix an inviolable law to himself, never, by any temptation, to be induc'd to violate" the principles of justice.[5] The other part of Darwall's argument for the claim that agents take the rules of justice to be prescriptive is conceptual. He argues that only by regarding the rules as prescriptive do agents have sufficient grounds for the regulation of their own behavior. Regarding the rules of justice non-prescriptively amounts to making a predictive judgment that others will regulate their conduct if one does also. This predictive judgment alone does not provide an agent with a ground (a reason or a motive) for conforming herself to these rules. The just agent, then, regards the rules of justice as prescriptive and normative in themselves: these rules themselves provide conclusive reasons for regulating one's behavior as specified by the rules.[6]

This suggests that the will of the just person aims at the performance of certain acts because they are required by the rules of justice. If this is Hume's view, then it seems to depart from his account of the will. By asserting this, Darwall is arguing that the will of the just person is not aiming at any good (pleasure) or at any avoidance of evil (pain), since what it is directed towards is simply the conformity with a rule. It will be helpful to note that Darwall takes Hume's view of the will to be orthodox in that it says that the will "invariably aims at the good," but he adds that Hume understands this good "hedonistically and egoistically."[7] Agents can set themselves to following a rule on this account, but only because doing so brings them pleasure (or allows them to avoid pain). This is inconsistent with the view that just agents see the rules of justice as prescriptive, normative, and providing conclusive reasons *by themselves* for acting.

---

[5] Darwall 1993, 422. I will provide the references to the *Treatise* when I will reconsider these passages below.

[6] Ibid.

[7] Darwall 1993, 423.

What's interesting about Darwall's discussion of Hume's theory of the will is that he seems to recognize that the theory he attributes to Hume is in tension with some other of Hume's commitments *besides* those arising from his discussion of justice. In chapter 5 above I briefly discussed the passions that Hume says "produce good and evil, and proceed not from them, like the other affections" (T 2.3.9.8, SBN 439). These include a desire for revenge against our foes, a desire to benefit our friends and family, as well as hunger, thirst, and our sexual appetites. Either these passions are consistent with Hume's theory of the will or they are not. If they are, then, in the case of revenge against my enemies, the will aims at the pleasure that exacting revenge would bring me. If the passions are not consistent with Hume's theory of the will, then the will may aim directly at the act of revenge itself. Darwall thinks Hume intends the latter. The will aims directly to obtain the objects of desire, or to remove the objects of aversion. Given that Hume grants the existence of, what Darwall calls, "nonstandard" direct passions, Hume is "committed to revising his theory of action and will quite independently of anything he says about the just person's regulation of her conduct."[8]

There is one further puzzle that arises from the relationship between justice and Hume's theory of the will as Darwall sees it. The will is "good-regarding" in a way that justice is not. Unlike other virtues, justice commands adherence even in cases where no good to oneself or others seems to arise. The rules of justice require that I repay my debts even when doing so will seriously impede my ability to live but where my doing so or not makes no difference to the well-being of my avaricious creditor. It is true that everyone, or most everyone, following the strictures of justice results in a common good, but it is not always the case that this common good is at all apparent to me. Hume says that, "the consequences of every breach of equity seem to lie very remote, and are not able to counter-ballance [*sic*] any immediate advantage, that may be reap'd from it" (T 3.2.7.3, SBN 535). This is especially so when the population of a society grows very large. I lose the connection between my own interests and my following the common

---

[8] Ibid., 424.

good when I start to see that my own isolated violations of the rules of justice are not very likely to upset the social order. Nevertheless, Hume makes it clear that just conduct is still expected: "When society becomes sufficiently numerous, individuals may tend to lose sight of their real interests, but these continue to dictate abiding by the rules of justice in every case" (T 3.2.2.24, SBN 499). So, Darwall concludes, "unlike benevolence and self-love, the trait of justice *itself* aims at no good to self or others, but at regulation by mutually advantageous rules (even though moral sentiment and self-interest both endorse the trait for its usefulness)."[9]

What Motive is Approved of in Just Acts?

The foregoing suggests that self-interest cannot serve as the motive to justice since justice is stable in ways that self-interested consequences are not.[10] Hume does claim that we have a *natural obligation* to the establishment of the conventions required for justice. Hume argues that in the state of nature, which is admittedly less hostile than Hobbes conceives it, self-love, a public regard, and a limited natural benevolence are unable to serve as motives for justices. However, he thinks that a "corrected" self-interest is a natural obligation to establish the practice of justice (T 3.2.1.10, SBN 480). This sort of self-interest takes into account the greater long-term goods that come about as a result of engaging in the mutually beneficial practices of justice.

However, the remarks at T 3.2.7.3 and T 3.2.2.24 make it implausible to suppose that corrected self-interest serves as the motive for just acts since Hume there seems to be conceding that there are cases where, in established and populous societies, even one's long-term good is secured by acting contrary to the requirements of justice.[11] Hume clearly rejects this as the motive for just acts in the following example:

---

[9] Darwall 1993, 419.

[10] I roughly follow Tate's presentation of Darwall's arguments in this section. See Tate 2005, 101-105.

[11] Tate draws attention to Hume's discussion of the "sensible knave" in the second *Enquiry* in order to show that Hume seems to think that genuine cases of free-riding are possible. See Tate 2005, 104; EPM 9.22-23, SBN 282-3).

> I suppose a person to have lent me a sum of money, on condition that it be restor'd in a few days; and also suppose, that after the expiration of the term agreed on, he demands the sum: I ask, *What reason or motive have I to restore the money?* It will, perhaps, be said that my regard to justice, and abhorrence of villainy and knavery, are sufficient reasons for me, if I have the least grain of honesty, or sense of duty and obligation. (T 3.2.1.9, SBN 479)

Hume later says that "we have naturally no real or universal motive for observing the laws of equity, but the very equity and merit of that observance" (T 3.2.1.17, SBN 483). But Hume's Circle, as we saw above, relies on the Undoubted Maxim: "*that no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the sense of its morality*" (T 3.2.1.7, SBN 479). This claim strongly suggests that agents cannot be motivated merely by their regard for the justice of their actions, since the virtue of justice depends on some other motive besides this regard.

Darwall calls the Undoubted Maxim the "fundamental thesis of his [Hume's] virtue ethics."[12] He claims that in order for us to take it to apply to all virtues (both natural and artificial), and in order for us to take his claims about the proper motive to just acts being a regard to equity itself, we have to allow for a view that is inconsistent with Hume's theory of the will. Just agents take the rules of justice to be normative and authoritative in themselves. They are motived by their regard for the regulation of their conduct by these rules since these rules provide them with conclusive grounds (motives or reasons) for acting. The virtue of the just agent is acting from a sense of their being under a rule obligation to act. This obligation does not involve an agent seeking any particular good for herself or for others. So, if this is the virtue of the just person, then it is not consistent with Hume's theory of the will.

The Trouble with Darwall's Hume

Margaret Tate points to one of the consequences of reading Hume as Darwall suggests. Recall the distinction between the two traditions Darwall locates in the British Moralists. On the one hand there are the empirical naturalists, who share a central commitment to explaining

---

[12] Darwall 1993, 429.

morality in purely naturalistic terms and rely on a conception of reason as a purely theoretical faculty. On the other hand, there are the autonomous internalists. I have said very little of this tradition but I think what I have said will suffice to show one important consequence of Darwall's discussion. The autonomist camp seeks to understand obligation in terms of "conclusive motives raised through the exercise of autonomous practical reasoning."[13] Darwall thinks that Kant develops a version of autonomist internalism, and so sees the roots of Kantianism in some of the British Moralists.[14] If we accept that Hume appeals to the notion of a rule obligation, and that he understands the virtue of justice as an "agent state" whereby agents regulate themselves by normative principles they accept as being intrinsically binding, then it seems that Hume is developing a theory that takes some significant steps away from the empirical naturalist tradition and moves closer towards the autonomist tradition. Darwall says, "even Hume feels the attraction of a more robust notion of normativity than any that can be supplied by a reductionist or reforming naturalism – at least, by those of his time."[15]

Given that "no philosopher has been more widely proclaimed than Hume as the source of the instrumental theory of rational action, as well, of the idea that normativity can be theorized as instrumental rationality," Darwall's interpretation requires a significant rethinking of our typical understanding of Hume. But this rethinking will have consequences for contemporary moral philosophy as well.[16] As Tate points out, one central concern of philosophers trying to provide an account of normativity is the proper conception of practical rationality. Given Hume's reputation as the champion of the empirical naturalist cause, showing that he himself recognized

---

[13] Darwall 1995, 16-17.

[14] Ibid., 323.

[15] Ibid., 321.

[16] Ibid., 320.

the limitations of this tradition gives "strong evidence, though not conclusive, for the inadequacy of empirical naturalism."[17]

The mere fact that Hume on Darwall's reading moves away from the intellectual tradition with which he has most commonly been associated is not, of course, in itself an objection to the interpretation. It may be that commentators were simply wrong in taking Hume to champion instrumental accounts of reason and naturalistically minded accounts of moral obligation. However, in order for this "new historiography" of Hume and his relation to Kant to be plausible, it should avoid, wherever possible, making Hume inconsistent. Discovering insights of historical figures that have been previously overlooked shouldn't come at the cost of rendering unintelligible other elements of their system, unless the textual and conceptual pressure is so great that such inconsistencies cannot be avoided. The problem with Darwall's reading of Hume is that it seems to depend on making Hume inconsistent, in rather glaring ways, and these inconsistencies seem to be entirely avoidable. In the following sections I will briefly provide responses to the puzzles that Darwall uses to motivate his interpretation. I do not pretend that I have solved all the difficulties that arise from Hume's conception of justice. What I hope to show is that there are plausible alternatives to Darwall's reading that do not result in glaring inconsistencies.

The Theory of the Will is More Complex

Hume's account of justice is supposedly inconsistent with his theory of the will because the will is "good-regarding" and justice is "rule-regarding."[18] Recall, though, that Darwall seems to think that Hume's theory of the will is inconsistent with his theory of the *productive* direct passions, like the desire for revenge, hunger, thirst and so on. Hume's remarks about the nature and operation of will do not unequivocally support hedonism and egoism, but this is an important

---

[17] Tate 2005, 94-5.

[18] Darwall 1993, 419.

supposition of Darwall's interpretation. We saw in chapter 4 that Hutcheson's remarks about the nature of desire and pleasure led to a similar sort of problem, namely that hedonism is suggested in places but is hard to reconcile with other claims. But Darwall's Hutcheson is a hedonist, and he likely thinks Hume just adopts this position.

But no clear commitment to hedonism or egoism emerges from the comments that Darwall cites in his discussion of the will. For instance, Hume says that "the *chief* spring and moving principle" of actions is pleasure and pain, and echoes this later by saying, "the *chief* spring or actuating principle of the human mind is pleasure or pain, and when these sensations are remov'd, both from our thought and feeling, we are, *in great measure*, incapable of passion or action, of desire or volition" (T 1.3.10.2, SBN 118; T 3.3.1.2, SBN 574; emphasis added). Neither of these say unequivocally that there is no motive, action, or volition without pleasure and pain. If we have reason to think that Hume claims some motives, actions, or volitions can arise independent of an agent's concern for pleasure or pain, then I think there is good reason to *not* read the passages as committing Hume to a kind of hedonism. The productive direct passions provide clear cases where Hume thinks this occurs. While it's true that I come to derive a great deal of pleasure from exacting revenge on my enemies, the initial desire I have for such revenge is not directed at this pleasure.[19]

Nor should we think, as Darwall seems to, that these are aberrant or "nonstandard" passions. Hume says that "beside good and evil, or in other words, pain and pleasure, *the direct passions frequently arise* from a natural impulse or instinct, which is perfectly unaccountable" (T 2.3.9.8, SBN 439; emphasis added). Darwall thinks that Hume is committed to revising his theory of the will in light of these direct passions. But, as Tate asks, "Is it appropriate to call observations made directly after the statements that Darwall accepts as Hume's theory of the will

[19] Tate suggests something like this view in 2005, 98.

a *revision* of that theory? Perhaps the theory was never so simplistic."[20] It is therefore unclear that Hume's theory of the will is solely good-regarding in the way that Darwall supposes.

Hume Solves the Circle

As I mentioned in the previous chapter, both Hume's Circle and the Undoubted Maxim are part of Hume's larger argument for the claim that justice is an artificial virtue. But Darwall treats the Undoubted Maxim as the fundamental principle of Hume's virtue ethics. This would itself be no problem if there were in Darwall a clear discussion of how the Undoubted Maxim were supposed to function in the Circle to establish the claim that justice is an artificial virtue. Of course, Hume's own discussion of this argument is less than clear. But this is no excuse to not attempt an interpretation of the argument, and Darwall, while he does pull quotes from the relevant passages, never attempts to provide an interpretation of Hume's argument for the claim that justice is an artificial virtue.

I think that Tate is right to point out that what this allows Darwall to do is to treat Hume's Circle as if "it were a dangling problem for Hume – a sign of the confusion that rule obligation and the agent state of rule regulation are supposed to clarify."[21] Darwall claims that Hume's views on justice are "paradoxical" and that he "confusedly asserts that it is the moral sentiment that uniquely motivates artificially virtuous acts. And that is what lands *him in the circle*."[22] Darwall thinks that Hume has identified a deep-seated problem with the virtue of justice and that Hume needs to abandon other commitments of his view in order to resolve it. But it is less than clear that Hume was worried that he had stuck himself with a paradox and that he'd need to change his own view in order to fix it. The problem that Hume's Circle presents is, apparently,

---

[20] Ibid., 97.

[21] Ibid., 106.

[22] Darwall 1995, 290; 315; emphasis added. I follow Tate in highlighting these passages since they show quite clearly the interpretive approach that Darwall takes towards Hume's Circle, namely, as a problem *for* Hume rather than an argument employed by Hume to show that the virtue of justice cannot be natural.

resolved by claiming that justice is an artificial virtue. Hume is explicit about this in the conclusion of the argument:

> From all this it follows, that we have naturally no real or universal motive for observing the laws of equity, but the very equity and merit of that observance; and as no action can be equitable or meritorious, where it cannot arise from some separate motive, there is here an evident sophistry and reasoning in a circle. Unless, therefore, we will allow, that nature has establish'd a sophistry, and render'd it necessary and unavoidable, we must allow, that the sense of justice and injustice is not deriv'd from nature, but arises artificially, tho' necessarily from education, and human conventions. (T 3.2.1.17, SBN 483)

Of course, it is open to Darwall to respond to this point by agreeing with Tate that Hume takes the artificial virtues to solve the problem of the circle. He would then point out that, on his interpretation, the artificial virtues involve rule obligations and self-regulating agency. So, what allows Hume to escape his own circle is to develop an account of obligation and agency that he saw was lacking in his predecessors. But it's not clear that the circle presents a problem for Hume's theory. It seems to present a problem for Hutcheson's account insofar as Hutcheson does not recognize any distinction between natural and artificial virtues, but instead understands all virtues in terms of benevolence. Hume doesn't think benevolence could motivate just acts in the state of nature. And he thinks that our natural limited benevolence would only move us occasionally to act justly in society. He doesn't think we'll be able to locate any natural motive for acting justly and that the only motive available is that such actions are just, equitable, fair and so on. Since these motives do not arise naturally, we need to suppose that they arise only on the introduction of some artifice or contrivance, a social convention. This is all that we get from the conclusion of the argument at T 3.2.1.17. While it may be true that Hume goes on to develop the account of the artificial virtues in the way that Darwall suggests, this development does not come within Hume's Circle itself, as Darwall seems to suggest.

I cannot here develop a complete interpretation of the artificial virtues or of the structure of Hume's Circle.[23] Hume thinks that the artificial virtues develop over time by the introduction

[23] I do think that something like Tate's interpretation is plausible. She suggests that the motives to justice develop gradually along with the virtue of justice itself. See Tate 2005, 106-116.

of self-interested conventions that gradually garner the approval of our moral sense. The account is sophisticated and nuanced and I cannot hope to do it justice here. But I do not think that I need to present my own interpretation of Hume's view here in order to continue my objection to Darwall. Darwall is the one who owes us an account of Hume's Circle. He is the one who needs to explain why Hume is not at all flustered when he finds himself stuck with a paradox. Given that Darwall sees Hume as not only rejecting the empirical naturalist tradition, but also rejecting other elements of his own view, I think that the burden of proof is both sufficiently high and falls squarely on Darwall's shoulders. It is so high, in fact, that I think we are safe in saying that Darwall fails to meet it merely because he is silent about the structure of the Circle.

What Else Is There?

There is a tendency in the history of British moral philosophy to read earlier figures as developing nascent theories that are eventually fully developed by later and greater thinkers. This tendency is especially strong when the later and greater thinker interpreters have in mind is Immanuel Kant.[24] This general tendency is not always a problem, since it is clear that there are some fairly stable developments of particular lines of thought running through British moral philosophy. For instance, it seems fairly clear that Richard Price attempts to develop a rationalist account that takes Clarke's view as its jumping-off point, but that is better able to handle the charges that Hutcheson and others had leveled against Clarke. But these lines are often difficult to trace and can remain rather uncertain. Beyond this, there is a real risk that we will read earlier figures as grappling with the same concerns their successors grapple with. Kant was clearly unhappy with naturalistic and reductive accounts of morality. If we read the British Moralists with an eye towards Kant, it seems only natural that we'll be ready to spot earlier figures who are

[24] Schneewind's account of the history of modern moral philosophy is called *The Invention of Autonomy* and seeks to recount modern moral philosophy as a long, winding path up to Kant. Even someone like Schneewind, who does not accept Kant' views, is still strongly tempted to look for the historical roots of Kant in earlier thinkers.

also unhappy with the empirical naturalist approach. Darwall just happens to make the startling, and textually unwarranted, suggestion that one such figure is an empirical naturalist himself.

But something else emerges when we read these figures on their own terms and don't try to saddle them with someone else's tack. We find that each figure provides a thoroughly naturalist account of moral obligation, and that they are capable of developing responses to different versions of the amoralist's challenge. While some might see these responses as being unsatisfactory, these are the views we're presented with nonetheless. I think that one of the difficulties that one encounters in understanding historical figures is that there is often a temptation to read one's aversion to a theory into the theory itself. We think, "that view cannot be correct" and then attempt to locate a more palatable alternative in the text provided. While something like this may be required by a principle of charity, I think there is a temptation to rely on it too heavily. I think that the foregoing discussion about Darwall's Hume suggests that he makes this sort of an interpretive move. In what follows, I argue that he also does so with respect to Hutcheson's view. There is, then, something very puzzling about Darwall's discussion of the sentimentalist version of empirical naturalism. However, Darwall's reading of Hobbes does not seem to be plagued by the same kind of difficulty. Even though it is clear that he thinks that Hobbes's view is problematic, he nevertheless provides a fairly compelling interpretation that falls rather neatly within the taxonomy of views I proposed at the outset. I briefly suggest an explanation for this difference.

Hobbes and Bindingness

In some ways, Hobbes is the empirical naturalist that is most amenable to Darwall's project of discovering an answer to the question, "in what does an *ought to do* consist?" We can see Hobbes's interest in developing an account of the "normative force" of moral obligations when he says, for instance, that, "Contracts oblige us, laws tie us fast being obliged," and then elaborates "a man is obliged by his contracts, that is, that he ought to perform for his promise sake; but that the law ties him being obliged, that is to say, it compels him to make good his

promise, for fear of the punishment appointed by the law" (DCv xiv.ii, R 100-1). Recall that Hobbes distinguishes between two related senses in which it is true that someone ought to do some act. There is a sense of obligation that derives analytically from one's having laid down a right in contracting. In this sense an agent $ought_O$ to keep her agreements. There is also the sense of obligation that arises from the law of nature, the rational *dictate* that one perform those acts necessary for her self-preservation. It is true of an agent that she $ought_L$ to keep her agreements when doing so is necessary for her self-preservation. The fact that they are related to one another, show that Hobbes was trying to account for the "force" that attends our moral judgments. An agent $ought_O$ to keep her agreement only if she $ought_L$ to keep her agreement. The "force" or "bindingness" that attends her keeping her agreements is result of the "force" or "bindingness" of her desire for self-preservation.

However, we must be clear that the kinds of judgments that Hobbes is recognizing as being "normative" are still always just judgments about what is likely in one's own interest. Darwall suggests that Hobbes at times talks as if there is this set of interdefinable, and irreducibly normative, terms.[25] But, and Darwall correctly notes this, Hobbes does not have room in his system for any such conceptions. When Hobbes says that we ought to keep our promise "for our promise sake" he is not attributing to promises any mysterious power, nor is he saying that we see certain rules as in themselves authoritative and binding. Hobbes thinks that keeping one's promises is a rule that, if followed, tends to bring about what is most in one's interests. This is one place where Darwall expresses dissatisfaction with Hobbes's instrumentalist empirical naturalism. He says, "it is far from clear that covenant will even be generally possible unless agents have available some motivational resource for keeping covenant other than self-preservation or self-interest, specifically, the capacity to follow a rule."[26]

---

[25] The normativity that Darwall is suggesting here would be such that it is "additional to anything science might discover about human beings, their passions, and their situation vis-à-vis another." (1995, 54).

[26] Ibid., 79.

One of the reasons that Hobbes's account of obligation is so amenable to Darwall's project, and why Hobbes's view so readily fits into Darwall's taxonomy, is that Hobbes is attempting to provide a naturalistic account of obligation that tries, in some way, to relate all senses of the term 'obligation' to what he calls 'material obligation'. Material obligation, again, is just being bound by external impediments. Hobbes does not reduce all obligations to material obligations. The literal "bonds" of material obligations seem to inspire Hobbes to think that other sorts of obligation can also be understood in terms of natural or supernatural "forces." He thinks the idea that an agent is, in some sense, unable to resist the power of a superior force (either some physical force, or God, or the sovereign) is crucial to its being the case that an agent is under an obligation. Finding a theory of the "bindingness" of ought judgments within Hobbes seems plausible because Hobbes took himself to be accounting for obligations in a way that relates them to literal bonds. It is not enough, Hobbes thinks, for agents to make verbal agreements with one another. They must also see making these agreements as in some sense binding them to their word.

It is also important to notice that the concept of natural law, natural right, justice, and obligation are central to Hobbes's moral philosophy. Hobbes's takes over much of the language of his natural law predecessors, and attempts to revise this language in such a way that coheres better with his metaphysical and epistemological commitments.[27] These concepts, while they do figure into Hutcheson's and Hume's accounts, do not have the same centrality.

Hutcheson's Anti-Juridical Approach

Hutcheson, however, seems less interested in accounting for anything like the "bindingness" of obligations. His sentimentalist account of obligation makes moral obligation a matter of the approbation of spectators. He insists that all humans have a moral sense as well as benevolent desires, and he argues at length that the benevolent desires are fundamental

---

[27] Schneewind 1998, 82-100.

determinations of human nature to the same extent that self-interested desires are. In understanding moral obligation in these sentimentalist terms, Hutcheson is moving away from a conception of morality where obligations, duties, and laws play a central role. He is instead more interested in describing the psychological makeup of moral agents and showing that, given this makeup, the motives approved of by the moral sense are standardly operative in us. We are already other-interested and -directed in our affections and our actions.

Of course, this still leaves room for conflict between our motivational states. Hutcheson does seem to allow that it is at least possible for there to be genuine conflict between self-interest and benevolence. Ordinarily, my greatest self-interest is best secured through my acting so as to promote the common good. This result is, for Hutcheson, secured by God since God has arranged things in such a way that my own greatest happiness is best sought out by seeking the happiness of others for its own sake, and I am already naturally driven to this end. However, it is not entirely clear that Hutcheson thinks we already have a natural benevolence towards the greatest system, or species and it is this latter which is most approved of by the moral sense. So, Hutcheson also seems to recognize a possible conflict between one's limited natural degrees of benevolence and one's benevolence towards the greatest number of sentient beings. Now even though I have argued that Hutcheson does have a response to the question "why ought I to expand my circle of concern," what this response shows is, again, Hutcheson's focus on the natural psychological propensities that human beings already have. The pleasures of reflective self-approbation ground a motivating reason for me to expand my circle of concern.

Recall that Darwall thinks that Hutcheson adopts the sentimentalist account of obligation without any evident philosophical rationale. Hutcheson accepts the view of practical reason that Darwall calls the calm reflective deliberation view. This view maintains that reason, understood as an entirely theoretical faculty, is capable of revealing new objects of our desires through its calm and reflective use. For Hutcheson, as we have seen, the calm reflective use of reason gives rise to, or reveals, our desire for the greatest happiness of the greatest number. However, Hutcheson does not accept the calm reflective deliberation version of *internalism*, which holds

that certain choices are morally obligatory because the calm reflective use of reason reveals to agents deliberating about what to do new desires for ends.[28] Instead, Hutcheson draws on his influence from Shaftesbury to provide a sentimentalist account of moral obligation, which holds that something's being morally obligatory amounts to nothing other than it giving rise in suitably specified spectators to a particular sort of sentiment.

I do not find Hutcheson's adoption of this account to be at all puzzling. In fact, I think the place to begin to see Hutcheson's philosophical rationale for adopting it is with one of Darwall's descriptions of sentimentalism: "For neither Hutcheson nor Hume does ethics fundamentally concern *regulation of conduct*, not to mention *self*-regulating moral agency."[29] If we return to Hutcheson's account of obligation in the *Inquiry*, we can see clearly that he was concerned to set his theory of obligation apart from his egoist opponents by how he frames his discussion of obligation: "If any one ask, can we have any sense of *obligation* abstracting from the *laws* of a *superior*?" (*Inquiry* VII.i, R 346). Neither sense of obligation depends on the existence of laws.[30] What is more, the moral sense of obligation doesn't depend on any sort of self-interested motive, and, by the time of the *Illustrations*, doesn't depend on any motive whatsoever. Hutcheson may be thinking that the Hobbesian is unduly concerned with legalistic concepts like law, rule, and regulation. His sentimentalism arises out of a conviction that, not only do juridical moral theories ignore natural benevolent motives, but they yield an ethical theory that places too much stress on governing and dictating behavior. Hutcheson's philosophical rationale for moving away from understanding moral obligation in terms of the motives that are raised through the use of calm reflective deliberation just is that such a view is still too focused on regulation of behavior.

---

[28] Darwall 1995, 106-7.

[29] Darwall 1995, 16. Darwall adds a footnote explaining that Hume's theory of justice is an exception to this claim.

[30] Notice that Darwall's argument that Hume must have adopted a third sense of obligation, rule obligation, depends on Hume's adopting Hutcheson's two conceptions of obligation.

Darwall represents Hutcheson as having a choice: he could understand moral obligation in terms of a motive (desire for the greatest good of the greatest number, say) or he could understand moral obligation in terms of a state of approval that is contingently and causally related to a motive. Hutcheson opts for the latter. However, note that he is able to incorporate something of the spirit of the former into his own view. If he had chosen the former view, it is likely, Darwall suggests, that he would adopt Cumberland's view that we're "obligated by the good of all."[31] This seems to mean that calm reflective deliberation reveals to us as our end the good of all, and we are morally obligated to act on those desires that arise from this sort of deliberation. But on Hutcheson's own account, to describe an act as being morally obligatory is to talk about states of approval directed at agent's motives. What the moral sense approves of most just is that motive that would have been given rise to through the calm reflective use of reason. Once we recognize this, I think that Hutcheson's reasons for taking the sentimentalist line become fairly clear. Moral obligations are not simply certain sorts of motives arrived at through particular processes of reason, moral obligations are moral precisely because of the states of approval of that sense which is the very foundation of morality. Now, if we continue to ask, "what makes the actions approved of by the moral sense obligatory," as Darwall seems to, I think Hutcheson will just say that his account can go no farther. They're obligatory because they are approved.

I argued above that Hutcheson was concerned with establishing the practicality of ethics. I argued that he thinks that where his view succeeds on this front, and where the egoist and the rationalist fail, is in his consistent championing of the greatest natural good for the greatest number. He thinks that, since this is the greatest good available, having it in constant view will give rise in us to a desire for it. He doesn't think that he needs to provide reasons for us to care for this end, even though his system does allow for such reasons. By tying moral obligation to a state of approval, rather than a motive, Hutcheson is able to say that we are naturally driven to

---

[31] Ibid., 234.

fulfill our moral obligations and given that the obligations are contingently connected with our states of approval, there is not much (though there is some) room for the problem of amoralism.

Internalism and Empirical Naturalism

Darwall's taxonomy of internalist positions does not map neatly onto the theories that are developed by Hobbes, Hutcheson, and Hume, but the problems are more serious for his discussion of the sentimentalists. I do think that Darwall correctly identifies a set of shared commitments among these theories regarding naturalism and the nature of reason. Hobbes and Hume provide very clear examples of theorists who attempt to explain morality without relying on any supernatural facts. And while it requires some additional comment to make this case for Hutcheson, I think that Darwall is right that he too shares this commitment. Hutcheson and Hume provide very clear examples of theorists who maintain that reason is a purely theoretical faculty insofar as they hold that reason is only able to discover truth and falsity. There is good textual reason to attribute this commitment to Hobbes as well. However, attributing this view to Hobbes is controversial and cannot be done without comment, but I do think it is the right reading of Hobbes. I think, then, that Darwall is right with respect to his claim that Hobbes, Hutcheson, and Hume all fall within the empirical naturalist tradition.

What is problematic about Darwall's discussion of this tradition is that he often speaks of it as a tradition of internalist theories. For example, he introduces the distinction between the empirical naturalist tradition and the autonomist tradition in *The British Moralists and the Internal 'Ought'* in a section titled "Two Internalist Traditions."[32] Yet Darwall thinks that neither Hutcheson nor Hume are strictly speaking internalists. This does not escape Darwall's notice. He says of Hutcheson's and Hume's sentimentalist accounts, "in addition to these approaches [those of Hobbes, Cumberland, and *Hutcheson*], another empirical naturalist theory of obligation was developed by British moralists that is internalist in a broad sense, although not

---

[32] Darwall 1995, 14.

in the specific sense of our study."[33] Darwall here reiterates his interpretation of Hutcheson as straddling two different accounts of obligation: the calm reflective deliberation view and sentimentalism. On Darwall's own reading, Hutcheson does not fall neatly in one or the other of these theories. Darwall thinks this is because of what he regards as an entirely unfounded move that Hutcheson makes in defining obligation in terms of a moral sense. But, it seems, this choice is only puzzling if we were already expecting Hutcheson to adopt Cumberland's view. But why should we expect this? Defining obligation in terms of rational motives may have been the going trend in moral philosophy at the time, as Darwall suggests, but this isn't a reason for us to expect Hutcheson to be on the bandwagon. Philosophers do go against current trends. Darwall will also likely point to Hutcheson's brand of practical reason has having affinities with Cumberland's view. But again, this is no reason alone to expect that Hutcheson would adopt Cumberland's account of obligation.

Indeed, when we look at Hutcheson's remarks on obligation in context, it becomes clear that his interest in obligation extends only so far as his defense of sentimentalism requires. In neither the *Inquiry* nor the *Illustrations* does obligation receive a substantial treatment. The *Inquiry* introduces the concepts of obligation only in the final section, where Hutcheson is concerned to show that the moral sense can account for "some complex moral ideas" (*Inquiry* VII). He defines the two senses of obligation and their relationship to our states of approval, and then quickly moves on to consider the role of education, and the nature of rights.[34] Hutcheson does talk about the aforementioned distinction between obligation and constraint in this section. But my main interpretive point remains: Hutcheson's discussion of obligation in the *Inquiry* arises from his attempt to corroborate his moral sense theory by showing how more complex moral ideas are given rise to through the operation of the moral sense.

---

[33] Ibid., 16.

[34] The nature of rights does not clearly depend on an account of obligation. See *Inquiry* VII vi-xii, R 353-354.

The *Illustrations* account of obligation also arises within the context of Hutcheson's arguments for sentimentalism. The account of obligation arises as a part of the Section I argument that all exciting reasons presuppose instinct and all justifying reasons presuppose a moral sense. He says, "let these Words, *Duty*, *Obligation*, *Owing*, and the meaning of that Gerund, *is to be preferred*, be explained; and we shall find our selves still at a Loss for *exciting Reasons* previously to *Affections*, or *justifying Reasons* without recourse to a *moral Sense*" (IMS 146). His primary interest in appealing to notions of obligation in the *Illustrations* is to show that one cannot make sense of the reasonableness of an action, independent of appealing to affections or a moral sense, by saying that the action is obligatory. Talking of moral obligation does not give an independent account of the reasonableness of an action.

Given this, then, I do not find Hutcheson's adoption of the sentimentalist account of obligation to be at all puzzling. Hutcheson is only concerned to address the notion of obligation in order to show that the moral sense theory has explanatory power, in the case of the *Inquiry*, and to show that rationalists cannot rely on the concept of obligation to establish an account of the reasonableness of action, in the *Illustrations* case. Given the larger argumentative contexts in which Hutcheson's discussions of obligation arise, it would be puzzling indeed if he did not adopt a sentimentalist account. Darwall seems to think that Hutcheson's sentimentalist account of obligation is arbitrary, and that he would have been somehow on better philosophical ground if he would have accepted a calm reflective deliberation version of internalism. But this isn't Hutcheson's view.

Darwall also claims that Hume accepts a sentimentalist account of obligation, and that he does so for equally seemingly arbitrary reasons. Both Hutcheson's and Hume's views fall within empirical naturalism even if they are not versions of empirical naturalist internalism, in a strict sense. However, Darwall does think they are broadly internalist accounts that hold that "moral obligation is internal."[35] I am not entirely sure what this claim amounts to. We saw that, for

---

[35] Darwall 1995, 16.

Hutcheson, there is a very strong causal connection between moral approbation and motivation. This causal connection is not a necessary connection, strictly speaking, but for beings with our constitution, Hutcheson seems to think it is safe to regard it as such. Maybe Darwall thinks something similar is happening in both Hutcheson and Hume. Or perhaps he thinks that the moral 'ought' is *internal* to observers rather being *internal* to agents as they are on strict internalist views. This is what Darwall says about Hutcheson's view not being strictly internalist.[36] But Darwall spends very little time making sense of Hume's *sentimentalist* account of moral obligation. His main focus is on developing his interpretation of Hume's virtue of justice. What's interesting about this interpretation is that, if it succeeds, there is a sense in which Hume would accept a strict version of internalism after all. Darwall's account of rule obligation is internalist in the sense that being rule-obligated to Φ means that one accepts as regulating for her own behavior some rule dictating that Φ is to be done. This is an autonomist internalist view since a rule obligation is a motive that is raised through the autonomous use of reason. In any case, when Darwall says that Hume's view, like Hutcheson's, is "internal," he does not mean to be including in this Hume's account of rule obligation. He admits this exception in a footnote.

So, on Darwall's view, here is what we can say about the empirical naturalist sentimentalist accounts of obligation: Hutcheson should not have held this theory, but he should have adopted Cumberland's view instead; and Hume should not have adopted this theory either, but he should have adopted something closer to Shaftesbury's or Butler's theory instead. So, it seems, classical sentimentalism with respect to moral obligation is simply a mistake! I think this result is not textually well-grounded. I think, for the reasons I specified above, that we are better off thinking about the sentimentalists as offering anti-juridical views where notions of moral obligation take a back to seat to naturally motivating sentiments and the psychological mechanisms by which they are approved. On Darwall's taxonomy, the sentimentalists seem to slide between different kinds of theories. I prefer to think that this suggests a limitation in the

---

[36] Ibid., 234.

"new historiography" than to think that is shows an instability in either Hutcheson's or Hume's view.

There is much richness and complexity in the moral theories of Hobbes, Hutcheson, and Hume and I hope that I have allowed some of that richness and complexity to come out in my discussions. I must admit that I have raised more interpretive issues than I have been able to settle, but this is no mark against me. My aim here has been to examine the early modern empirical naturalists in order to discover what they say about the relationship between obligation, motivation, and reasons for acting. In particular, I have been trying to discover in what sense such theories could be internalist moral theories and, what, if anything, they can say to address the question "why be moral" (this is, after all, the "modern moral problem"). Asking whether or not Hobbes, Hutcheson or Hume would have accepted internalism is a bad question given the myriad ways in which philosophers understand 'internalism.' So we must ask for clarification on what it takes to be an internalist and what it takes to be an externalist. Hobbes, Hutcheson, and Hume are silent on this question, but I think we can excuse them for this. So, we are left developing our own taxonomies and using these as our guide. My suggestion here is that Darwall's taxonomy does not, to borrow the metaphysician's expression, carve empirical naturalist theories of obligation at the joints.

REFERENCES

Árdal, Páll S. 1966. *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press.

Baier, Annette C. 1991. *A Progress of Sentiments: Reflections on Hume's "Treatise"*. Cambridge, MA: Harvard University Press.

Blackstone, William T. 1965. *Francis Hutcheson and Contemporary Ethical Theory.* Athens: University of Georgia Press.

Bricke, John. 1996. *Mind and Morality: An Examination of Hume's Moral Psychology.* Oxford: Clarendon.

Brown, Charlotte. 1988. "Is Hume an Internalist?" *Journal of the History of Philosophy* 26 (1): 69-87. doi:10.1353/hph.1988.0021.

Capaldi, Nicholas. 1989. *Hume's Place in Moral Philosophy*. New York: Peter Lang.

Cohon, Rachel. 1988. "Hume and Humeanism in Ethics." *Pacific Philosophical Quarterly* 69 (2): 99-116.

———. 2008. *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press.

———. 2010. "Hume's Moral Philosophy." *The Stanford Encyclopedia of Philosophy.* http://plato.stanford.edu/archives/fall2010/entries/hume-moral/.

Coleman, Dorothy. 1992. "Hume's Internalism." *Hume Studies* 18 (2): 331-47. doi:10.1353/hms.2011.0421.

Darwall, Stephen. 1983. *Impartial Reason*. London: Cornell University Press.

———. 1989. "Motive and Obligation in the British Moralists." *Social Philosophy and Policy* 7 (1): 133-150.

———. 1992 "Internalism and Agency." *Nous-Supplement: Philosophical Perspectives* 6:155-174. http://www.jstor.org/stable/2214243.

———. 1993. "Motive and Obligation in Hume's Ethics." *Noûs* 27 (4): 415-448. http://www.jstor.org/stable/2215786.

———. 1995. *British Moralists and the Internal 'Ought' 1640-1740.* Cambridge: Cambridge University Press.

———. 1997. "Hutcheson on Practical Reason." *Hume Studies* 23 (1): 73-90.

———. 2000. "Normativity and Projection in Hobbes's Leviathan." *Philosophical Review* 109 (3): 313-47.

Fabian, Bernhard. 1971. Preface and Bibliography note to *An Inquiry in the Original of Our Ideas of Beauty and Virtue*. In *Collected Works of Francis Hutcheson* vol. 1 prepared by Bernhard Fabian. Hildesheim: Georg Olms.

Falk, W. D. 1948. "Ought and Motivation." *Proceedings of the Aristotelian Society* 48: 111-38. http://www.jstor.org/stable/4544438.

Feinberg, Joel. 1977. "Wollaston and His Critics." *Journal of the History of Ideas* 38 (2): 345-352. doi:10.2307/2708918

Foot, Philippa. "Hume on Moral Judgment" In *Virtues and Vices: and Other Essays on Moral Philosophy*, 74-80. Oxford Scholarship Online. doi: 10.1093/0199252866.001.0001.

Frankena, William. 1955. "Hutcheson's Moral Sense Theory." *Journal of the History of Ideas* 16 (3): 356-375. http://www.jstor.org/stable/2707637.

———. 1970. "Obligation and Motivation in Recent Moral Philosophy." In *Readings in Ethical Theory*, 2nd ed. Edited by Wilfred Sellars and John Hospers, 708-729. New York: Appleton-Century-Croft.

Fumerton, Richard. 1990. *Reason and Morality: A Defense of the Egocentric Perspective.* Ithaca: Cornell University Press.

Garrett, Don. 2006. "Hume's Naturalistic Theory of Representation." *Synthese* 152 (3): 301–319. http://www.jstor.org/stable/27653398.

Gauthier David. 1979. "Thomas Hobbes: Moral Theorist." *The Journal of Philosophy* 76 (10), Seventy-sixth Annual Meeting of the American Philosophical Association, Eastern Division: 547-559. http://www.jstor.org/stable/2025550.

———. 1982. "Three Against Justice: The Foole, The Sensible Knave, and the Lydian Shepherd." *Midwest Studies in Philosophy*. 7 (1): 11-29. doi:10.1111/j.1475-4975.1982.tb00081.x.

———. 1997. "Hobbes on Demonstration and Construction." *Journal of the History of Philosophy* 35 (4): 509-521. doi:10.1353/hph.1997.0080.

Gert, Bernard. 1967. "Hobbes and Psychological Egoism." *Journal of the History of Ideas* 28 (4) (1967): 503-520. http://www.jstor.org/stable/2708526

———. 1991. Introduction to *Man and Citizen (De Homine and De Cive)*, by Thomas Hobbes, 3-32. Edited by Bernard Gert. Indianapolis: Hackett Publishing Company.

———. 1996. "Hobbes's Psychology" In *The Cambridge Companion to Hobbes*, edited by Tom Sorell, 157-174. Cambridge: Cambridge University Press. doi:http://dx.doi.org/10.1017/CCOL0521410193.008.

———. 2001. "Hobbes on Reason." *Pacific Philosophical Quarterly* 82 (3-4): 243-257. doi:10.1111/1468-0114.00127.

Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition.* Cambridge: Cambridge University Press.

———. "Feminist Contractarianism" In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by Louise M. Antony and Charlotte Witt, 227-255. Boulder: Westview Press.

Harrison, Jonathan. 1976. *Hume's Moral Epistemology.* Oxford: Clarendon Press.

Hayes, Peter. 1999. “Hobbes's Silent Fool: A Response to Hoekstra”. *Political Theory* 27 (2): 225–229. doi:10.1177/0090591799027002004.

Hobbes, Thomas. 1969. *De Corpore* in *The English Works of Thomas Hobbes of Malmesbury*. vol. 1 Edited by William Molesworth. London: John Bohn. First published 1655.

———. 1991. *Man and Citizen (De Homine and De Cive)*. Edited by Bernard Gert. Indianapolis: Hackett Publishing Company. English translation of *De Cive* first published 1651. *De Homine* first published 1658.

———. 1994. *Leviathan: With Selected Variants from the Latin Edition of 1668*. Edited by Edwin Curley. Indianapolis: Hackett Publishing Company. First published 1651.

Hoekstra, Kinch. 1997. "Hobbes and the Foole." *Political Theory* 25 (5): 620-654. doi:10.1177/0090591797025005002.

———. 1999. "Nothing to Declare? Hobbes and the Advocate of Injustice." *Political Theory* 27 (2): 230-235. doi:10.1177/0090591799027002005.

Hume, David. 1998. *An Enquiry Concerning the Principles of Morals*. Edited by Tom L. Beauchamp. Oxford: Oxford University Press. First published 1751.

———. 1999. *An Enquiry Concerning Human Understanding*. Edited by Tom L. Beauchamp. Oxford: Oxford University Press. First published 1748.

———. 2009. *A Treatise of Human Nature*, revised ed. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press. Books 1 and 2 first published 1739. Book 3 first published 1740.

———. (1755) 2008. *Selected Essays*. Edited by Stephen Copley and Andrew Edgar. Oxford: Oxford University Press.

Hutcheson, Francis. 2002. *An Essay of the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense.* Edited by Aaron Garrett. Indianapolis: Liberty Fund Inc. First published 1728.

———. 2008. *An Inquiry in the Original of Our Ideas of Beauty and Virtue*, revised edition. Edited by Wolfgang Leidhold. First published 1725.

Jeske, Diane. 2008. *Rationality and Moral Theory: How Intimacy Generates Reasons.* New York: Routledge.

Jensen, Henning. 1971. *Motivation and the Moral Sense in Francis Hutcheson's Ethical Theory.* The Hague: Martinus Nijhoff.

Joynton, Olin. 1981. "Wollaston's Theory of Declarative Actions." *Southern Journal of Philosophy* 19 (4): 439-450. doi:10.1111/j.2041-6962.1981.tb01448.x.

Kalt, Stefan. 2005. "David Hume on the Motivating Effect of Moral Perception: Internalist or Externalist?" *History of Philosophy Quarterly* 22 (2): 143-160. http://www.jstor.org/stable/27745019.

Kavka, Gregory S. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.

———. 1995. "The Rationality of Rule-Following: Hobbes's Dispute with the Foole." *Law and Philosophy* 14 (1): 5-34. http://www.jstor.org/stable/3504931.

Kendrick, Nancy. 2009. "Why Hume's Counterexample is Insignificant and Why It is Not." *British Journal for the History of Philosophy* 17 (5): 955-979. doi:10.1080/09608780903135113.

Korsgaard, Christine. 1986. "Skepticism about Practical Reason." In *Foundations of Ethics: An Anthology*, edited by Russ Shafer-Landau and Terence Cuneo, 299-312. Malden: Blackwell Publishing Ltd.

———. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Locke, John. 1979. *An Essay Concerning Human Understanding*. Edited by Peter Nidditch. Oxford: Clarendon Press. First published 1751. First published 1690.

Loeb, Louis E. 1977. "Hume's Moral Sentiments and the Structure of the Treatise." *Journal of the History of Philosophy* 15 (4): 395-403.

Lowe, E. Jonathan. 2005. *Locke: The Routledge Philosophers*. Abingdon: Routledge.

Mackie, J. L. 1980. *Hume's Moral Theory.* London: Routledge & Kegan Paul Ltd.

Martinich, A.P. 1992. *The Two Gods of* Leviathan*: Thomas Hobbes on Religion and Politics*. Cambridge: Cambridge University Press.

Mautner, Thomas, ed. 1993. Introduction to *Francis Hutcheson: On Human Nature with Reflections on Our Common Systems of Morality and On the Social Nature of Man.* Cambridge: Cambridge University Press.

McIntyre, Jane L. 2000. "Hume's Passions: Direct and Indirect." *Hume Studies* 26 (1): 77-86. http://www.humesociety.org/hs/issues/v26n1/mcintyre/mcintyrev26n1.pdf.

Moore, G.E. (1903) 1988. *Principia Ethica*. Buffalo: Prometheus Books.

Murphy, Mark. 1994. "Deviant Senses of 'Obligation' in Hobbes' *Leviathan*." *History of Philosophy Quarterly* 11 (3): 281-294. http://www.jstor.org/stable/27744629.

Norton, David F. 1985. "Hutcheson's Moral Realism." *Journal of the History of Philosophy* 23 (3): 397-418. doi:10.1353/hph.1985.0046.

———. 2009. Introduction and annotations to *A Treatise of Human Nature*, revised ed., by David Hume, I9-I107. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.

Owen, David. 2009. "Hume and the Mechanics of Mind: Impressions, Ideas, and Association." In *The Cambridge Companion to Hume*, 2nd ed. Edited by David Fate Norton and Jacqueline Taylor, 70-104. Cambridge: Cambridge University Press.

Paletta, Douglas R. 2011. "Francis Hutcheson: Why Be Moral?" *Journal of Scottish Philosophy* 9 (2): 149-159. http://dx.doi.org/10.3366/jsp.2011.0013.

Parry, Geraint. 1967. “Performative Utterances and Obligation in Hobbes.” *The Philosophical Quarterly* 17 (68): 246-252. doi:10.2307/2218158.

Paxman, Katharina. 2015. "Imperceptible Impressions and Disorder in the Soul: A Characterization of the Distinction between Calm and Violent Passions in Hume." *Journal of Scottish Philosophy* 13 (3): 265-278. doi:10.3366/jsp.2015.0108.

Peach, Bernard. 1954. "Methods of Analysis in Ethics: Francis Hutcheson and W. K. Frankena on Obligation and Ability." *Methodos* 6: 287-98.

———. 1971. Introduction to *Illustrations on the Moral Sense* by Francis Hutcheson. Cambridge, MA: Belknap Press.

Peacock, Mark. 2010. “Obligation and Advantage in Hobbes’ *Leviathan*.” *Canadian Journal of Philosophy* 40 (3): 433-458. http://www.jstor.org/stable/25763348.

Penelhum, Terence. 2009. “Hume's Moral Psychology.” In *The Cambridge Companion to Hume*, 2nd ed. Edited by David Fate Norton and Jacqueline Taylor, 238-269. Cambridge: Cambridge University Press.

———. 2015. "The Indirect Passions, Myself, and Others." In *The Cambridge Companion to Hume's Treatise*. Edited by Donald C. Ainslie and Annemarie Butler, 206-229. New York: Cambridge University Press. http://dx.doi.org/10.1017/CCO9781139016100.011

Prichard, Harold Arthur. 1949. *Moral Obligation; Essays and Lectures.* Oxford: Clarendon Press.

Radcliffe, Elizabeth S. 1994. "Hume on Motivating Sentiments, the General Point of View, and the Inculcation of 'Morality'." *Hume Studies* 20 (1): 37-58. doi:10.1353/hms.2011.0000.

———. 1996. "How Does the Humean Sense of Duty Motivate?" *Journal of the History of Philosophy* 34 (3): 383-407. doi:10.1353/hph.1996.0058.

———. 1999. "Hume on the Generation of Motives: Why Beliefs Alone Never Motivate." *Hume Studies* 25 (1-2): 101-122.

———. 2004. "Love and Benevolence in Hutcheson's and Hume's Theories of the Passions." *British Journal for the History of Philosophy* 12 (4): 631-653. doi:10.1080/0960878042000279305.

———. 2006. "Moral Internalism and Moral Cognitivism in Hume's Metaethics." *Synthese* 152 (3): 353-370. doi:10.1007/s11229-006-9003-6.

———. 2013. "Moral Sentimentalism and the Reasonableness of Being Good." *Revue Internationale de Philosophie* 1 (2013): 9-27.

Raphael, D. D. 1947. *The Moral Sense*. London: Oxford University Press.

———,ed. 1991. *British Moralists 1650-1800.* 2 Vols. Indianapolis: Hackett Publishing Company.

Rhodes, Rosamond. 1992. “Hobbes's unReasonable Fool.” *Southern Journal of Philosophy* 30 (2): 93-102. doi:10.1111/j.2041-6962.1992.tb01717.x.

———. 2002. "Obligation and Assent in Hobbes's Moral Philosophy." *Hobbes Studies* 15 (1): 45-67. doi:10.1163/187502502X00046.

Ross, W. David. (1930) 2002. *The Right and the Good*. Edited by Phillip Stratton-Lake. Oxford: Clarendon Press.

Russell, Bertrand. 1970. "The Elements of Ethics." In *Readings in Ethical Theory*, 2nd ed. Edited by Wilfred Sellars and John Hospers, 3-28. New York: Appleton-Century-Croft.

Schauber, Nancy. 1999. "Hume on Moral Motivation: It's Almost Like Being in Love." *History of Philosophy Quarterly* 16 (3): 341-366. http://www.jstor.org/stable/27744825.

Schneewind, J. B. 1998. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.

———. 1990. *Moral Philosophy from Montaigne to Kant: An Anthology*. 2 vols. Cambridge: Cambridge University Press.

Scott, W.R. 1900. *Francis Hutcheson: His Life, Teaching and Position in the History of Philosophy*. Cambridge: Cambridge University Press.

Smith, Michael. 2007. "The Externalist Challenge." In *Foundations of Ethics: An Anthology*, edited by Russ Shafer-Landau and Terence Cuneo, 231-242. Malden: Blackwell Publishing Ltd.

Smith, Norman Kemp. 1941. *The Philosophy of David Hume: A Critical Study of Its Origins and Central Doctrines.* N.p.: Palgrave Macmillan, 1941.

Sorell, Tom. "Hobbes's Scheme of the Sciences." In *The Cambridge Companion to Hobbes*, edited by Tom Sorell, 45-61. Cambridge: Cambridge University Press. doi:http://dx.doi.org/10.1017/CCOL0521410193.003.

Stevenson, C.L. [1937] 1970. "The Emotive Meaning of Ethical Terms." In *Readings in Ethical Theory*, 2nd ed. Edited by Wilfred Sellars and John Hospers, 254-266. New York: Appleton-Century-Croft.

Strasser, Mark. 1990. *Francis Hutcheson's Moral Theory: Its Form and Utility.* Wakefield: Longwood Academic.

Tate, Margaret Watkins. 2005. "Obligation, Justice, and the Will in Hume's Moral Philosophy." *Hume Studies* 31 (1): 93-122. doi:10.1353/hms.2011.0202.

Taylor, A. E. 1938. "The Ethical Doctrine of Hobbes". *Philosophy* 13 (52): 406–24. http://www.jstor.org/stable/3746387.

Tenenbaum, Sergio. 2000. "Ethical Internalism and Glaucon's Question." *Noûs* 34 (1): 108-130. http://www.jstor.org/stable/2671975.

Tweyman, Stanley. 1976. "Truth, Happiness and Obligation: The Moral Philosophy of William Wollaston". *Philosophy* 51 (195): 35–46. http://www.jstor.org/stable/3749767.

Warrender, Howard. 1957. *The Political Philosophy of Hobbes: His Theory of Obligation.* London: Clarendon Press.

Wedgwood, Ralph. 2009. "The ‘Good’ and the ‘Right’ Revisited." *Philosophical Perspectives* 23 (1): 499-519. doi:10.1111/j.1520-8583.2009.00181.x.

Weller, Cass. 2002. "The Myth of Original Existence." *Hume Studies* 28 (2): 195-230.

———. 2004. “Scratched Fingers, Ruined Lives, and Acknowledged Lesser Goods.” *Hume Studies* 30 (1): 51-85. doi:10.1353/hms.2011.0328.

Winkler, Kenneth P. 1985. "Hutcheson's Alleged Realism." *Journal of the History of Philosophy* 23 (2): 179-194. doi:10.1353/hph.1985.0032.