

12-13-2013

# Toward a Naturalistic Solution to the Moral Problem

David J. Lambie

*University of Connecticut - Storrs*, david.lambie@uconn.edu

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

---

## Recommended Citation

Lambie, David J., "Toward a Naturalistic Solution to the Moral Problem" (2013). *Doctoral Dissertations*. 300.  
<https://opencommons.uconn.edu/dissertations/300>

## Toward a Naturalistic Solution to the Moral Problem

David J. Lambie, PhD

University of Connecticut, 2013

The purpose of my dissertation is to work toward providing a naturalistic solution to the moral problem. The moral problem arises because there are three theses relevant to morality which are independently plausible but collectively implausible. First, moral statements are ones that appear to have objective truth values. Second, if a moral statement is true then it appears that, other things being equal, we ought to do as the statement says. Third, all behaviors are caused by desires—that is, the Humean theory of motivation is true. Michael Smith develops a solution to the moral problem which, if successful, would reconcile the tension that exists between the three theses. Much of the work necessary for completing this task involves providing a plausible account of the connection between moral judgment and moral motivation.

The first chapter of my dissertation lays out some of the main views that have been held concerning the connection between moral judgments and moral motivation. The rest of the dissertation consists of two main projects. The first involves arguing that Smith's solution is implausible. In chapter two I draw upon pre-existing work as well as my own original contributions to argue that Smith's case against moral motivational externalism falls short. In chapter three I argue that Smith's solution to the moral problem does not deliver substantive moral truths. The second half of the dissertation focuses on developing an Aristotelian-style virtue ethics solution to the moral problem. Chapter four focuses on moral motivation in Aristotle and argues that his account need not be interpreted as internalist. In addition, an independent argument is offered in support of motivational externalism. In chapter five I defend virtue ethics from some criticisms that could be advanced against it and offer an analogy with mechanical systems to help dispel one particularly damning criticism. I defend the analogy itself from a number of objections that could be advanced against it.

# Toward a Naturalistic Solution to the Moral Problem

David J. Lambie

B.A., University of Minnesota, 2002

M.A., University of Connecticut, 2008

A Dissertation

Submitted in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy  
at the  
University of Connecticut

2013

Copyright by  
David J. Lambie

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

Toward a Naturalistic Solution to the Moral Problem

Presented by

David J. Lambie, B.A., M.A.

Major Advisor \_\_\_\_\_  
Thomas D. Bontly

Associate Advisor \_\_\_\_\_  
Samuel C. Wheeler III

Associate Advisor \_\_\_\_\_  
Michael P. Lynch

University of Connecticut  
2013

## Acknowledgements

I want to express my appreciation to my advisor, Thomas Bontly, for his patience and valuable feedback on numerous drafts of the chapters within this dissertation as well as feedback on previous drafts of my dissertation prospectus. Without his support this work would not have been possible. I also want to express my appreciation for my committee members Sam Wheeler and Michael Lynch for their feedback on previous work as well as their support throughout this challenging process.

This dissertation was written while I was working as an adjunct professor at the State University of New York at Oswego. I want to express my sincere appreciation for the support and advice offered by the entire department. I especially want to thank K. Brad Wray, Craig DeLancey, Rob Card and Lori Nash for their support and good advice on all sorts of matters, both practical and professional. I also want to thank my good friend and office mate Richard Cocks for his support, invitations to Thanksgiving dinner, and for challenging my analytic philosophical perspective from time to time.

I wish to thank all my friends and family for their support throughout the years. Most importantly, I wish to thank my parents, Sharon and Wallace Lambie, for their unwavering love and support, without which I would not have been able to make it as far as I have. I dedicate this work to them.

## Table of Contents

Introduction	1
Chapter One	4
1.1 Introduction	4
1.2 Social Normativity	6
1.3 Psychological Normativity 1	9
1.4 Psychological Normativity 2	17
1.5 The Skeptical Objection and Response	22
1.6 Conclusion	25
Chapter Two	27
2.1 Introduction	27
2.2 The Case for Moral Externalism	27
2.3 Critiquing the Color Analogy	34
2.4 Critiquing the Political Example	38
2.5 Critiquing the Fetishism Argument	44
2.6 Alternate Explanations of the Close Connection between Moral Judgment and Moral Motivation	49
2.7 Summary and Preview	54
Chapter Three	55
3.1 Introduction	55
3.2 Smith's Solution to the Moral Problem	55
3.3 The Method of Reflective Equilibrium	65
3.4 Smith's Argument against Besires	75
3.5 Evaluating Smith's Argument against Besires	78
3.6 Interlude on the Nature of Mental States and Our Justification for Believing in them	86
3.7 Summary and Preview of the Next Chapter	90
Chapter Four	92
4.1 Introduction	92
4.2 Internalism, Externalism and Dual-State Motivation	92
4.3 The Role of Reason and Desire in Virtuous and Morally Continent Persons	96
4.4 Desire and Reason	99
4.5 Objection 1: The Lack of Phronesis in the Morally (In)continent	101
4.6 Objection 2: The Directedness of Desire	102
4.7 Aristotle's Account of Moral Education	109
4.8 Conclusion	113

Chapter Five	114
5.1 Summary of Our Progress So Far and the Project for This Chapter	114
5.2 The Moral Problem Revisited	116
5.3 Objection 1 and Response—Three Theories of Besires	118
5.4 Objection 2 and Response—A Virtue Ethics Approach	124
5.5 Objection 3 and Response—An Aristotelian Justification of Moral Objectivity	129
5.6 Objection 4 and Response—The Mechanical Systems Analogy	138
5.7 Five Objections to the Mechanical Systems Analogy with Responses	155
5.8 Summary	168
Bibliography	170



## Introduction

The realm of morality presents a challenge to those who are interested in understanding its nature. Moral statements appear to express facts that are truth evaluable. We can capture this in a T-schema. 'Stealing is wrong' is true if and only if stealing is wrong. It is natural to think that acts of stealing must have a property of wrongness in order for the statement to be true. A number of questions can be asked here. For example, we might ask to what the property of wrongness would amount. Unfortunately, the T-schema does not provide much help here. It is not clear that science does either. Wrongness does not appear to be a physical property whose nature science can help us to understand.

Another puzzling aspect of morality is its particularly close connection with human action. Most facts seem to motivate only in conjunction with other factors. Suppose I am walking by a yard sale and see on display a book on the joys of bottle cap collecting. Whether my recognition of the fact that there is such a book motivates me to buy the book depends upon a number of particular factors about myself such as whether I enjoy bottle cap collecting, whether I enjoy collecting books on collecting, whether I am easily swayed to try new things, whether I can afford to buy anything, and so on. There is some evidence that moral facts are different. If I recognize that stealing is wrong I ought to feel at least some motivation not to steal. If I am not at all motivated to refrain from stealing, people are likely to judge that there is something wrong with me. Such a judgment is not likely if I feel no motivation to buy the book on bottle cap collecting.

The particularly close connection between what appear to be moral facts and our motivations to act in accord with those facts presents a difficult problem for the moral realist. Why should any fact about the way things are be motivating independent of any other beliefs and desires we happen to possess? This may present an even larger problem for the moral realist who is also a naturalist as opposed to the realist who embraces some form of rationalism. If we adopt a rationalistic strategy we can try to explain the connection between moral facts and

motivation by explaining right and wrong in terms of what we would do if we were rational. Our beliefs about what our most rational self would do might then serve to create desires to act in this way, though this is not an uncontroversial claim. The naturalist seems to face a more daunting challenge. If moral facts are natural facts just like any other facts then the link between the recognition of a moral fact and any motivation that accompanies it should depend on other beliefs and desires that we possess.

The purpose of this dissertation is to work toward providing a naturalistic solution to what Michael Smith calls the moral problem which he discusses in his book of the same name. The moral problem arises because there are three theses relevant to morality which are independently plausible but collectively implausible. First, moral statements are ones that appear to have objective truth values. Second, if a moral statement is true then it appears that, other things being equal, we ought to do as the statement says. Third, all behaviors are caused, at least in part, by desires—that is, the Humean theory of motivation is true. Smith develops a solution to the moral problem which, if successful, would reconcile the tension that exists between the three theses. Much of the work necessary for completing this task involves providing a plausible account of the connection between moral judgment and moral motivation.

The first chapter of my dissertation lays out some of the main views that have been held concerning the connection between moral judgments and moral motivation. The rest of the dissertation consists of two main projects. The first, which occupies chapters two and three, involves arguing that Smith's solution is implausible. In chapter two, I draw upon pre-existing work as well as my own original contributions to argue that Smith's case against moral motivational externalism falls short. In chapter three I argue that Smith's solution to the moral problem does not deliver substantive moral truths. The second half of the dissertation focuses on developing an Aristotelian-style virtue ethics solution to the moral problem. Chapter four focuses on moral motivation in Aristotle and argues that his account need not be interpreted as internalist. In addition, an independent argument is offered in support of a type of external

motivation. In chapter five, I defend virtue ethics from some criticisms that could be advanced against it and offer an analogy with mechanical systems to help dispel one particularly damning criticism. I finish up by defending the analogy itself from a number of objections that could be raised to it.

## Chapter One

### Section 1.1 Introduction

Moral discourse seems to contain both factual and normative elements. Our linguistic practices treat utterances such as ‘Killing the innocent just for the fun of it is morally bad’ and ‘Giving to charity when one can afford it is the right thing to do’ as purporting to express facts which have truth values. (Most moral reasoners would judge both to be true.) Nor do we judge someone who has made either statement as suffering from less than a full understanding of our language practices. In addition to being factual, moral discourse seems to be normative as well. The judgment that a moral statement is true seems to dictate that we behave in appropriate ways. The truth of (e.g.) ‘Killing the innocent just for the fun of it is morally bad’ seems to require that we refrain from killing the innocent just for the fun of it, that we try to get others to refrain from killing the innocent just for the fun of it, and that there be some sort of consequences for people who do kill the innocent just for the fun of it.

This dualistic nature of moral discourse results in what Michael Smith calls *The Moral Problem* and is discussed in his book of the same name. The project of my dissertation will be to try to provide a naturalistic solution to this problem. According to Smith, the following three theses are individually plausible but collectively inconsistent.

1. Moral judgments of the form ‘It is right that I  $\phi$ ’ express a subject’s beliefs about an objective matter of fact, a fact about what it is right for her to do.
2. If someone judges that it is right that she  $\phi$ s then, *ceteris paribus*, she is motivated to  $\phi$ .
3. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume’s terms, distinct existences.<sup>1</sup>

The main problem here is that moral beliefs seem to necessarily have an action guiding nature that appears to be lacking in other beliefs. For example, imagine someone has just learned that

---

<sup>1</sup> Smith (1994), p. 12.

the Earth is approximating 93 million miles away from the Sun. There seems to be no necessary connection between this fact and any corresponding action.<sup>2</sup>

I will be assuming the truth of Michael Smith's first thesis because I assume the truth of moral realism. The focus of this dissertation will be on the second and third theses. The second thesis tells us that if someone makes a moral judgment about some factual matter then that person will be motivated to act in certain ways that are in accord with that judgment whereas the third concerns the arrangement of mental states that appears to be necessary to explain the connection between motivation and belief. The second and third theses concern the normative nature of morality. As a result, a discussion of moral normativity as it relates to the project of naturalistically-based moral realism will be helpful.

When put in general terms the normativity of morality is not difficult to understand. True moral beliefs do not just tell us about some state of the world but also have an action-guiding nature. If it is morally wrong to eat meat then, at the very least, those of us who do should try to stop. So we can understand moral normativity as the action-guiding nature of moral beliefs. It is also of little doubt that morality actually is normative. Knowing that moral normativity consists in an action-guiding property or properties and that morality must have it does not tell us much about the nature of these properties. As we dig deeper, things become more complicated. There seem to be three places these properties might be located: within human minds, within the natural world that exists outside of human minds, and in some non-natural source such as God. As this essay consists in an attempt to provide a naturalistic solution to the moral problem I will not explore any non-natural solutions. Therefore, I set aside the possibility that moral normativity is located in a non-natural source.

---

<sup>2</sup> While normative issues are present in areas other than morality, the linkage between the factual and the normative is particularly strong in the moral realm. I will not concern myself with non-moral normativity in this work.

Understanding the nature of moral normativity is essential for the moral naturalist because an inadequate account of normativity can be a source for moral skepticism. Richard Joyce expresses the concern in the following way:

The problem, however, and my main ground for doubting the (moral naturalist's) project, is that in order to naturalize moral clout we cannot be content just to find a property that has practical authority—arguably we have located such a property in *being-such-that-you-would-want-to-do-it-if-you-were-to-reason-correctly*. We must also satisfy inescapability; we need a property that has this authority over people *irrespective of their interests*. But it is doubtful that any naturalizable account can deliver this . . . What the moral naturalist evidently needs is a substantive and naturalizable account of “correct practical reasoning” (or “practical rationality”) according to which any person, irrespective of her starting desires, would through such reasoning converge on certain practical conclusions that are broadly in line with what we would expect of moral requirements.<sup>3</sup>

Joyce sets the bar high for the moral naturalist. In this chapter I will try to get clear on just what moral normativity could amount to for the naturalist and what the naturalist is and is not required to explain when it comes to normative phenomena. I divide normativity up into two broad areas, social normativity and psychological normativity.

### Section 1.2 Social Normativity<sup>4</sup>

Imagine that an intelligent being that was completely unfamiliar with morality was to study human morality. The observer would notice the normativity in many places. One place the observer would notice the normativity of morality is in the language we use. Moral language seems to have both a descriptive side as well as a more directly, action-guiding nature. In practice, we recognize that moral statements which appear to express facts such as *it is wrong to steal* entail ‘ought’ statements such as *one ought not to steal*. Such imperatives do not seem to follow from most other factual statements. The fact that a student will do poorly on an exam if he or she does not study for it does not entail that the student ought to study for the exam given

---

<sup>3</sup> Joyce 2006, p.196.

<sup>4</sup> For lack of a better term, I call the area in this section *social normativity*. It might more appropriately be called *normativity from a third person perspective*.

that the student does not care about his or her performance. This distinction in the relation between imperatives and moral and non-moral facts was pointed out by Immanuel Kant.

The normativity of morality is reflected in our laws as well. There is significant overlap between our laws and moral beliefs. Practices such as murder, robbery, rape and pedophilia are not just considered immoral but are punishable by law. When the law allows what we deem to be immoral behaviors, such as slavery, or punishes behaviors we do not consider to be immoral, such as prohibiting homosexual behavior, we seek to change the laws. The laws also provide for punishments for persons who break the law but punishing immoral behaviors extends beyond the bounds of the law. It is common to desire revenge on those whose immoral behavior has harmed us and sometimes we act on these desires. When the moral transgressions are less serious we may decide to avoid the other person. In addition to punishing wrongdoers, we reward those who do good things.

The observer would also notice that we teach morality to our children, both directly and by promoting the kinds of emotional responses that are likely to lead to moral behavior. If my child pushes another child down I might both inform my child that it is wrong to push people down and promote empathy by asking her how she would feel if she were in the other child's place. In the United States of America, moral instruction is common in both primary and secondary schools. Colleges commonly offer and require courses in both theory of ethics as well as courses on practical ethics like business ethics and medical ethics.

The observer might come to the conclusion that moral practices exist because they play an important role in the functioning of a group.<sup>5</sup> Human beings do live in large and complex groups and do so quite successfully. Such a conclusion gains further support when one notices that many moral norms appear to fall into one of two broad categories; there are norms that reduce or prevent harm and norms that facilitate helping behaviors. Prohibitions against

---

<sup>5</sup> Of course, the observer would want to consider other hypotheses as well, such as that morality benefits the individual, that morality has no function or that it is foisted upon the masses by powerful elites.

harmful behaviors facilitate successful group living in at least two ways. A member of the group who is injured will be less able to do his or her part to make the group function well and may become a burden to other members of the group because there is one less person who can defend the group from outside attacks, participate in cooperative ventures, and so on. Second, a member of the group who has been harmed or attacked or had someone they care about harmed or attacked may be likely to retaliate. Ongoing feuds within the group will be likely to lead to a weakening of the group as a whole. Second, we all find ourselves in situations where we need help, at least on occasion. If we are helped we will be more likely to be able to play a productive role in the group and be less likely to become a burden. If a moral transgression has taken place and someone has been harmed as a result, helping the harmed party can serve to repair the harm done.

There appears to be broad agreement among naturalistically-minded moral philosophers that the purpose of morality is to facilitate the success of group living. Allan Gibbard suggests that “the key to human moral nature . . . lies in *coordination* broadly conceived.”<sup>6</sup> Gibbard develops a moral theory in which the acceptance of norms plays a key role in moral judgments. Being in the same community we will tend to accept the norms that are present in that community. Having widespread, but not necessarily unanimous, agreement concerning the norms that we accept allows us to coordinate our behavior with others in our community.

David Copp presents a naturalistic theory he calls society-centered moral theory.<sup>7</sup> According to Copp, moral statements, as well as certain non-moral statements like statements of etiquette, express standards. Copp assumes that accepting a moral code is necessary for every society. Each society has a moral code which helps it to meet three basic needs: it helps to maintain the physical integrity of the society, it helps guarantee the internal social harmony necessary for cooperation, and it promotes peaceful relations with neighbors. The moral

---

<sup>6</sup> Gibbard 1990, p. 26.

<sup>7</sup> Copp 1995.



propositions which are true are the ones that would be included in the ideal moral code, i.e. the code that best meets the three basic needs.

Philip Kitcher tentatively offers a picture of morality that has its root in what we might call the proto-morality of chimpanzees.<sup>8</sup> Kitcher points out that a chimpanzee will often form alliances with one or a few other members of the group. However, these alliances are quite unstable and require an inordinate amount of grooming time to repair and maintain. (Chimpanzees spend much more time grooming each other than they need to for parasite control.) Kitcher speculates that our ability to live in large groups and take on numerous cooperative activities without spending hours a day grooming each other is due to our having evolved a capacity for normative guidance. Our ability to internalize behavioral norms or rules allows for cooperative activity and, in turn, successful group living on a large scale.

### Section 1.3 Psychological Normativity 1—The Nature of the Connection

Our alien observer would certainly be missing out on something important about moral normativity if he/she/it only paid attention to the social aspects of morality. Almost all of us possess a first person experience of moral normativity in addition to the evidence we see of morality in the areas discussed above. There clearly exists some sort of connection between accepting something as a moral fact and being motivated to act in certain ways. There appear to be two questions of primary importance. First, what is the nature of the connection between apparent moral facts and moral motivation? Second, how strong is the connection between the two? I will start with the first question.

That there is a connection between accepting some moral fact and being motivated to act in the appropriate way seems impossible to deny. If I think that stealing is wrong and am presented with an opportunity to steal, my belief that stealing is wrong is likely to have some inhibitory effect on my behavior, even if my temptation for the object of my desire is stronger

---

<sup>8</sup> Kitcher 2006.

and leads me to steal. Some see the connection between moral facts and moral motivation as being a necessary connection.<sup>9</sup> This means that a necessary condition for a sincere judgment that some act is morally right (wrong) must be accompanied by some motivation to (not) do X.

This necessary connection can be captured by the following two conditionals:

- (A judges that X is morally wrong  $\supset$  A is motivated to not do X)
- (A judges that X is morally right  $\supset$  A is motivated to do X)

One advantage of what are called non-cognitivist views in ethics is that some have the resources to explain a necessary connection between moral facts and moral motivation.<sup>10</sup> In brief, non-cognitivist views in ethics are ones that say that an utterance like ‘Stealing is morally wrong’ does not express a belief. That is, there is no moral fact, be it a fact about the world or a truth which can be arrived at through a priori reasoning for the utterance to accurately represent. So when someone utters ‘Stealing is morally wrong’ we should take the person to be expressing something other than a belief. Different varieties of non-cognitivist views do this in different ways.

A. J. Ayer’s emotivist theory of ethics claims that when we are making a claim such as, ‘Dishonesty is morally bad’, we are not attributing the moral property ‘badness’ to acts of dishonesty.<sup>11</sup> This is because, for Ayer, terms like (morally) wrong, right, bad or good, have no factual content because they are not reducible to empirical concepts.<sup>12</sup> However, ethical terms do have the purpose of expressing and arousing feelings as well as motivating action.<sup>13</sup> But whenever we are making a moral judgment, Ayer tells us, “the function of the relevant ethical

---

<sup>9</sup> This is sometimes called ‘motivational internalism’. I avoid discussions of internalism here but the subject will be discussed in subsequent chapters.

<sup>10</sup> A good source for the distinction between non-cognitivist and cognitivist views in ethics is McNaughton 1988.

<sup>11</sup> See Ayer 1936, ch. 6.

<sup>12</sup> Ayer distinguishes between the descriptive and the normative uses of these terms. If these terms are used descriptively then they can be reduced to empirical concepts. For example, if we made the statement ‘Premarital sex is wrong’ to show a certain group’s disapproval of premarital sex then we ‘wrong’ would be reducible to the attitudes of the members of the group.

<sup>13</sup> Ayer, p. 108.

word is purely 'emotive'. It is used to express feeling."<sup>14</sup> It becomes clear why we would tend to think of the connection between apparent expressions of moral facts and motivation as necessary if we adopt an emotivist approach to ethics. This is because there are no moral facts in the relevant sense that would correspond to the feeling. So the connection would be between negative or positive feelings and being motivated in certain ways. Arguably, having negative (positive) feelings about some sort of act in and of itself is sufficient for being motivated to act in a certain way, even though such a motivation might be overridden by other motivations one has.

According to R. M. Hare, the primary moral sense of the word 'good' is to recommend. The word can be used in some other ways that are what Hare calls an "inverted commas" use. We can use a word in this way when we are being ironic (e.g. Golfers are "athletes", Adolf Hitler was a "good" person) or if we are ascribing a property that something has relative to a class (Tiger Woods is an "athletic" golfer, this is a good golf club) where these statements can be taken as being short for 'Tiger Woods is athletic for a golfer', This is good as a golf club.<sup>15</sup> But these are not the primary uses of the word 'good'. For Hare, its primary function is to commend and the purpose of a commendation is always "at least indirectly, to guide choices, our own or other people's, now or in the future."<sup>16</sup> Moral words like 'wrong' or 'bad' have similar functions as does 'good' except that whereas 'good' is used to commend, these words are used to condemn. Regardless of whether the speaker says, 'You ought not to steal' or the apparently descriptive 'Stealing is morally wrong', what the speaker is really doing is recommending that the hearer not steal. In fact, if there is a conflict between what we say and what we do—suppose I say that stealing is morally wrong but steal—we must conclude that I do not sincerely believe that stealing is morally wrong. Thus, Hare seems to accept the even stronger thesis that akrasia is not possible, except in special cases in which we are physically or

---

<sup>14</sup> Ibid, p. 108

<sup>15</sup> See Hare 1952, pp. 124-125.

<sup>16</sup> Ibid, p.127.

psychologically unable to do that which we believe is best.<sup>17</sup> It is the action-guiding aspect of morality that is primary and there are no independently existing facts, even if our language sometimes misleads us into thinking that there are.

Gibbard advances a complex version of non-cognitivism.<sup>18</sup> The view starts with the observation that there are norms that govern peoples' behaviors and distinguishes between norms that we accept and norms that we internalize. Some norms that we accept are also ones that we have internalized but on some occasions the two do not coincide. Further, an internalized norm may be one we accept in one situation but would not accept in another. Rational behavior is then defined in terms of the acceptance of norms. A behavior is rational just when it is consistent with norms that we accept. Gibbard speculates that human beings may possess two kinds of motivational systems, a *normative control system* and an *animal control system*.<sup>19</sup> The normative control system is the more distinctly human motivational system in which we come to accept a norm through conversation or internal conversation (i.e. thinking). The animal control system is older and is one that we presumably share with other animals. In certain cases these motivational systems will conflict and when the older animal control system wins out we experience what we call weakness of the will.

Gibbard's norm-expressivism has the resources to explain instances of weakness of the will. As a result, Gibbard's view does not require that the behavioral norms that we accept be overriding.<sup>20</sup> It is not clear to me whether Gibbard's view requires the acceptance of a norm to always be accompanied by some degree of motivation. There is some textual evidence to suggest that it does. Gibbard tells us that in addition to being used to share beliefs, language evolved to allow us to plan and thereby coordinate our activities.<sup>21</sup> This could not have happened if language did not have the power to motivate. So it is reasonable to think that there

---

<sup>17</sup> See Hare 1963, pp.77-79.

<sup>18</sup> See Gibbard 1990.

<sup>19</sup> Ibid, p. 56.

<sup>20</sup> The claim that morality is "overriding" will be discussed in the next section.

<sup>21</sup> Gibbard 1990, p. 57.

will normally be some motivation which accompanies accepting a norm.<sup>22</sup> Gibbard compares normative governance to political governance and says that, like political governance, “normative governance will not always prevail, but it has some influence on what people do and feel.”<sup>23</sup> If the evolutionary purpose of norm acceptance is to make possible cooperative ventures then we should expect that it would normally be accompanied by motivation.

Whether we are expressing our acceptance of a norm, recommending a behavior, or expressing our emotional reaction to a behavior, it is clear that each of these theories has the resources to explain the connection between moral beliefs and moral actions in a forceful way. A necessary connection between moral beliefs and motivations may seem desirable—though this will be questioned shortly—but it is not clear that the connection actually is a necessary one.

It may be that some of these non-cognitivist theories make the connection between moral facts and motivation too strong. Michael Stocker has argued that the connection between moral beliefs (he uses the term ‘the good’) and motivation is not necessary but instead is “mediated by large arrays of complex psychic structures, such as mood, energy, and interest.”<sup>24</sup> Of course, it is possible that the perceived good usually motivates because it is normal for peoples’ psychic structures to be configured in such a way that we will be attracted to what we believe is good, though Stocker suggests that this may not be the case. He says that the psychology that tends to be presumed by philosophers is that of the “successful and striving . . . man.”<sup>25</sup>

---

<sup>22</sup> Whether this motivation will really be necessary or merely contingent seems to hinge on whether a norm can be accepted without the creation of motivation or remain an accepted norm without accompanying motivation. This raises a substantive question about the physiology of the human brain that will not be explored here.

<sup>23</sup> Gibbard 1990, p. 77.

<sup>24</sup> Stocker (1979), p. 739.

<sup>25</sup> Ibid, p. 753.

However, let us assume that it is the norm for people to be attracted to the good.<sup>26</sup> I want to focus on three examples that Stocker gives of people not being attracted to the good. First, various forms of depression seem to lead us to not be attracted to the good.<sup>27</sup> Frequently, depressed people cease to be concerned with pursuing what they recognize to be the good for themselves. A depressed person may engage in self-destructive behavior, such as using harmful drugs or refusing to take much needed medication, even though they recognize that this behavior is not good for them. There is a temptation to dismiss such behavior as one in which the person does not actually believe that refraining from drug use or taking one's medication is good because they see themselves as having no self-worth. Yet milder forms of this kind of behavior are quite common in otherwise normal people. We might be motivated to eat a food that we know is not good for us.<sup>28</sup> For example, a person with high cholesterol may know that she should choose a salad for dinner rather than pizza. She knows the pizza is not good for her but desires the enjoyment that she will receive from eating it. We may even stipulate that she is a moral philosopher who whole-heartedly rejects hedonistic theories of the good.

A third case Stocker presents is one of a politician who in the past worked very hard to alleviate people's suffering but who now is unmoved to help.<sup>29</sup> This could be for a variety of reasons. One reason Stocker suggests is that the politician may believe that he did enough in his youth. The politician's case though may be expanded to cover a common and interesting phenomenon.<sup>30</sup> Young people are often quite idealistic and have a strong desire to solve social problems like global warming, poverty, and various kinds of discrimination. As we age, many of us feel less and less motivated to solve such problems, while still recognizing that it would be good to solve them.

---

<sup>26</sup> For stylistic reasons, I drop 'perceive' from in front of the 'good'. It should be understood that what the person is attracted to, or fails to be attracted to, is the perceived good.

<sup>27</sup> Ibid, p. 744.

<sup>28</sup> Ibid, p. 747.

<sup>29</sup> Ibid 741-742.

<sup>30</sup> This phenomenon is common in the United States. I make no claims about its presence anywhere else.

Stocker has described various forms of amorality. The presence of such cases suggests that the connection between beliefs about the moral good and motivation is contingent. Theoretically speaking, should we prefer a necessary connection to a contingent one? David O. Brink argues that views that require a necessary connection between moral beliefs and motivation cannot explain amorality and that they ought to be able to do so.<sup>31</sup> This is because such views must maintain that the amoralist is conceptually impossible. The non-cognitivist must maintain the alleged amoralist who acknowledges that e.g. some outcome is good but feels no motivation to pursue it must be using 'good' in the inverted commas sense. That is, they are saying that the outcome would be good but they do not really mean it. Such a solution is unsatisfying because it rules out by fiat a problem that seems to be a real issue.<sup>32</sup>

There are at least two additional reasons why a contingent connection between moral beliefs and motivations is preferable to a necessary connection. If we were creating moral agents then it may be thought desirable to make the connection between the agents' moral beliefs and their motivations a necessary one. However, people do sometimes make serious mistakes about the good. For example, many Nazis appeared to seriously believe that it was morally good to rid the world of Jews. An unmotivated Nazi is certainly preferable to a motivated one. A necessary connection between the *actual* good and peoples' motivations would presumably be desirable but the existence of such a connection is highly implausible from a naturalistic perspective and that is the perspective we are adopting here.

Second, it is not clear what evidence could be brought to disprove the claim that the connection is necessary. That someone would say something like the following does not strike me as being at all unusual—"I know that it was morally wrong to take the money out of that wallet I found but I *really* wanted to buy an iPod." Clearly the person was motivated to buy an iPod. However, this does not show that person was *not* also motivated to return the wallet. The

---

<sup>31</sup> By amorality here I mean what Brink means by strong amorality. See Brink 1989, p. 48.

<sup>32</sup> Ibid, p. 84.

natural feeling of the use of the term 'really' in this context might seem to suggest that there was a motivation to do as morality demands but that the motivation to buy an iPod won out because it was especially strong. The question concerning the strength of moral demands will be examined in the next section. However, it is not clear that there really is a motivation to return the wallet in this case. In justifying our behavior we may feel the need to claim that our temptation for an iPod was really strong but this does not show that there clearly was a competing demand.

Further, even if there is a competing motivation it is not clear that it is the recognition of a moral consideration that is doing the motivating. The behavior may be explicable in terms of the social aspect of normativity. For example, the thief may recognize that doing morally bad things are the kinds of things that one gets punished for and her desire to not be punished is what is providing the competing motivation. This is not to deny that there are cases in which a moral motivation competes with a non-moral motivation but we should not assume that all cases are of this sort. If one thinks "I really want an iPod but people may wonder where I suddenly got the money to buy one, maybe they'll suspect that I must have been the one who found Jones' wallet, and then I'll get in trouble", the competing motivations are not necessarily moral motivations.

What about people who routinely break a moral rule? Some people habitually steal even though they know it is morally wrong. Must we assume that on each occasion, there is some moral motivation, regardless of how weak it is? For some it will be tempting to assume that there must be some moral motivation here but that it goes unnoticed. However, another plausible story is that the habitual thief might "get over" his early conflicting moral motivations so that through his habit of stealing he no longer feels any (moral) motivation to refrain from stealing. The real concern for the naturalistically-minded moral philosopher is likely to be the inclusion of a substantive claim which no evidence could disprove. This should be seen as a flaw rather than an advantage of a moral theory. It is not immediately apparent under which



conditions the claim that there is a necessary connection between moral facts and actions could be disproven. It may be that at some point in the future brain scans will allow us to map what we call beliefs about the good and their corresponding motivations. If these beliefs and motivations wind up being the same mental state this would seem to provide conclusive evidence for a “necessary connection” (identical things always coexist) whereas if they differ this would signal a contingent connection.

#### Section 1.4 Psychological Normativity 2: The Strength of the Connection

There exists among many a strong conviction that the motivation that accompanies the moral reasons that we accept must be stronger than any other type of motivation. When this is so, our moral reasons will override our non-moral reasons.<sup>33</sup> This apparent feature of morality presents a bit of a biological puzzle. On the one hand, we should expect that we will take moral reasons to be overriding. Suppose that they are not overriding. In such a case, when conflicts arise with non-moral reasons the non-moral reasons will sometimes override the moral reasons. Yet, if the purpose of morality is to make social living possible it seems that they must be perceived as non-optional in some sense by their possessors. This is especially important because it is precisely when there is a conflict between moral reasons and other reasons, like reasons of self-interest, that moral reasons need to be effective. Suppose I have moral reason not to kill people. Most of the time I feel no temptation to kill anyone so it doesn't really matter if I have a moral reason to not to kill people. It is precisely when I feel the temptation to kill someone—perhaps my neighbor has angered me greatly—that my moral reason should override the reason my self-centered desire has provided. Otherwise, what would be the point of possessing that reason? Thus, we might assume that whichever selection pressures

---

<sup>33</sup> Here I am referring to internal reasons. See Williams 1981. Reasons will be discussed in much more detail in the following two chapters. A particularly interesting question exists concerning the type of mental state a reason could be.

selected for these moral norms would also predispose their possessors to take them as non-optional.

On the other hand, there will be selection pressures in favor of not always treating moral reasons as overriding. Suppose that it is in my self-interest to steal something that belongs to my neighbor, whom I don't really care for anyway. I know that the likelihood of being found out is practically zero and what I would steal is something that could benefit me greatly. Stealing under these types of circumstances might have provided survival advantages to our ancestors. More generally, we would expect transgressions to occur when these transgressions would benefit the transgressor with little likelihood of getting caught. Over the next few pages I will look at the tendency to see morality as overriding and authoritative, though I will need to be selective. I will focus on Philippa Foot's criticisms of Kant moral philosophy.

In Kant's system, moral imperatives, or duties, are categorical in nature; they are not contingent upon desires that we have. Philippa Foot challenges Kant on this point.<sup>34</sup> The main distinction between categorical imperatives and hypothetical imperatives is that the latter depend upon a person's desires whereas the former do not. Foot points out that there are rules of etiquette that also seem to be categorical in that they do not fail to apply just because someone does not desire to comply with them.<sup>35</sup> It seems that there must be something more that is meant when we say that moral judgments are categorical. Foot considers the possibility that moral considerations differ from rules of etiquette in that the former but not the latter provide reasons to act.<sup>36</sup> The problem lies in spelling out in just what this *reason-giving force* really consists. If moral considerations do have reason-giving force then (assuming other areas lack it) this will make moral considerations overriding. Foot locates this reason-giving force not within morality itself but the way morality is taught and the feelings it arouses in us.<sup>37</sup>

---

<sup>34</sup> Foot 1972.

<sup>35</sup> Ibid, p. 308.

<sup>36</sup> Ibid, p. 309.

<sup>37</sup> Ibid, pp.310-311.

In a later paper, Foot questions whether moral considerations really are always overriding.<sup>38</sup> She distinguishes between two kinds of moral considerations which she calls evidential considerations and verdictive considerations.<sup>39</sup> Evidential moral considerations are instances of acts, such as stealing or promise-breaking, that we normally think of as having moral relevance. As an example of evidential moral considerations not always being overriding Foot cites the tendency for the host of a party to continue to serve alcohol to his or her guests, even after it is clear that someone has drunk more than they should. This is an example of considerations of etiquette overriding moral considerations.

Verdictive considerations involve judgments that some act has a particular moral property (i.e. is morally right or wrong).<sup>40</sup> Here we may actually explicitly acknowledge what the right thing to do is but say that we really have no choice and must do the wrong thing. For example, a boss may fire an employee, admitting that it is morally wrong, but claiming that she has no choice. If she does not do so, her own boss will come down and do the firing and will fire her as well. The person doing the firing may feel bad about what she does yet she fires the employee anyway, knowing full well that it is not morally right. Moral considerations are not always *actually* overriding in either evidential or the verdictive cases.

What then is the source of the belief that moral considerations are overriding whereas (e.g.) rules of etiquette are not? Foot points to a difference between the ways the two codes are taught. Exceptions to rules of morality tend to be incorporated within morality itself. It is okay to break a rule of etiquette in cases of emergency but the judgment that you have still broken a rule of etiquette remains. We can also think of morality as having rules. For example, there is a rule that says that you ought to keep your promises. Suppose you have promised to meet a friend for dinner. However, just before your dinner plans a medical emergency arises and you

---

<sup>38</sup> See Foot 2002

<sup>39</sup> Ibid p. 182.

<sup>40</sup> It is being assumed in both cases that we are talking about people who care about moral considerations.

do not keep your promise. Our most natural judgment of this case is that you did nothing wrong. You did not break a rule because the moral rule requiring a person to keep their promises regarding dinner dates has a built in exception for situations in which keeping the promise would result in serious harm.

We teach morality in such a way that moral rules are more important than other kinds of rules, such as the rules of etiquette. The stringency of moral teachings may then provide an explanation for our feelings that there is a “binding force”<sup>41</sup> that accompanies morality. If the purpose of morality is to make group living possible, as many naturalists assume, then it makes sense that morality would be taught in such a way that strong feelings accompany it.

The stringency with which morality is taught, the feelings that accompany it, and our tendency to incorporate exceptions to moral rules within the realm of morality provide the tools for a compelling explanation of why we would tend to view morality as overriding. Let’s look a little more closely at moral overridingness to see to what this claim is supposed to amount.

Consider the following two formulations of what might be meant by moral overridingness.

1. Moral motivations (reasons) will always actually override non-moral motivations (reasons).
2. Moral motivations (reasons) ought always to override non-moral motivations (reasons).

If (1) is what is meant by moral overridingness then the claim is shown to be false by empirical considerations, moral motivations (reasons) do not always override non-moral motivations (reasons). So (1) is clearly false. Foot suggests that (2) can be explained by the combination of our moral teaching and our feelings. The exact source of our conviction that (2) is correct is hard to pinpoint. It may be that feelings and moral teaching work in tandem, each reinforcing the other. Teaching our children that moral considerations must take precedence over other considerations is likely to result in our children believing (and feeling) that this must be so. Indeed, it seems that when a purported fact is stipulated rather than being arrived at because it is supported by evidence, it will be difficult to produce evidence for that purported fact. The

---

<sup>41</sup> Foot 1972, p. 310.

feelings generated from this teaching helps assure that these children will teach morality in the same way to their children. Additionally, if the moral naturalists who believe that the purpose of morality is to make group living possible are correct then it is plausible that we will have evolved in such a way that we are prone to react emotionally in moral situations.

None of this indicates that there is anything in the nature of morality itself that requires that moral motivations or reasons ought to be overriding. The common belief that moral reasons or motivations ought to be overriding can be explained in terms of social and evolutionary considerations and need not be intrinsic to morality itself.<sup>42</sup> From the previous section it became clear that a moral naturalist need not maintain that there is a necessary connection between moral beliefs and motivation. Empirical concerns based on Stocker's work give us reason to suspect that a moral belief about something's goodness can frequently exist without an accompanying motivation. If the advocate of a necessary connection claims that there must be a motivation even when there appears not to be, he risks supporting a theory that is not open to empirical refutation, regardless of the evidence that is presented. Brink argued that a contingent connection between our moral beliefs and motivation may be preferable from a theoretical standpoint as advocates of a necessary connection will have difficulty explaining the possibility of amorality.<sup>43</sup> Yet it is clear that there is a persistent and reliable connection between believing a moral fact and accompanying motivation and this connection is one that the moral naturalist should explain.

At the beginning of this section, I discuss a biological puzzle that the connection between overridingness and morality presents. Naturalistic theories seem well poised to

---

<sup>42</sup> It may be that when we have settled on the right moral theory we will see that overridingness is intrinsic to morality but this is not essential to the moral realist's project. As far as I can see, even if there is nothing in the nature of morality that explains why morality ought to override other considerations this does not mean that we should not teach that moral considerations ought to override other considerations. If a person's immoral behavior is going to negatively affect me, either because it harms me directly or because it weakens the social structure in which I live this makes it more important than breaches of etiquette and so on. Therefore, enlightened self-interest appears to dictate that we emphasize that, generally speaking, moral reasons ought to override non-moral ones.

<sup>43</sup> This point has also been raised by McNaughton 1988, ch. 9 though McNaughton is more optimistic about the possibility that the internalist can provide an acceptable explanation of the problem.

explain the apparent inconsistency between the fact that we typically think that moral reasons ought to override other reasons when they come into conflict and the fact that they do not actually do so a significant amount of the time. Earlier we discussed Gibbard's speculation that there might be two different kinds of motivational systems at work in human beings, an animal control system and a normative control system. If this is correct the two may compete for control within a person with the normative control system making a person profess one thing while the animal control system leads the person to do something else.<sup>44</sup> We may sincerely profess to accept one norm and, because we accept it, believe that it should override any other norms. There are also non-moral examples of this phenomenon. Suppose, you go to the eye doctor and the doctor tries to put drops in your eyes. However, you keep blinking even though you want to keep your eyes open.<sup>45</sup> This too can be interpreted as a conflict between two kinds of control system. One kind of control system has the purpose of keeping foreign objects out of the eye. The other kind of control system allows us to have conscious control over our behavior. The two can conflict with the result being that we intend to do one thing (keep our eyes open) but actually do the opposite.

### Section 1.5 The skeptical objection and response

Before concluding this chapter, it seems appropriate to discuss a skeptical objection to the findings thus far. The skeptic of the prospects of moral naturalism may respond to the discussion of normativity in this chapter as follows—so much the worse for moral naturalism. In a recent response to David Copp's attempt to provide a naturalistic form of moral realism<sup>46</sup>, Sharon Street says the following:

What (Copp's) theory tells us is what *moral* reasons we have; really what it speaks to is how *morally* to live. But if one wants to know how one has reason to

---

<sup>44</sup> Gibbard 1990, p. 57.

<sup>45</sup> The example is from Millikan, 2004, p. 3. Millikan's work will be discussed in greater detail in chapter five in an attempt to find a naturalistic solution to the moral problem.

<sup>46</sup> Copp 2008. See also Copp 1995.

live *simpliciter*—in other words, if one is asking how to live *period*—then (Copp’s) society-centered theory has nothing to say in answer to this question.<sup>47</sup>

Street is complaining that Copp’s view may tell us what kinds of moral reasons and that these reasons may exist along with other reasons like reasons of self-interest. The theory does not, however, tell us that we ought to choose the moral reason rather than the reason of self-interest should the two conflict. Street is echoing Joyce’s concerns in the passage quoted at the beginning of the chapter. Joyce says that the moral naturalist needs to provide a naturalized account of practical reason in which a person would decide that the right thing to do was in line with what we generally think of as the right thing to do.

I do not think that this objection is as strong as skeptics like Joyce and Street would like to believe. The key thing to keep in mind is that we are assuming that the moral naturalist has an otherwise defensible theory of morality. Given this assumption we are now concerned with what the normative requirements of such a theory would be. The skeptic is now arguing that it is not enough for the moral naturalist to have an explanation of moral normativity in which it is contingent and not always overriding. Consider the following thought experiment. Imagine that a completely novel philosophical theory is introduced which argues that the moral good is pleasure and the moral bad is pain. The theory generates a great deal of discussion among philosophers and is responsible for much research which lasts for many years. Early objections to the theory are dealt with convincingly and no serious counterexamples to the theory have been raised in a number of years. The theory is extended to cover almost all the major areas of our moral experience. For example, the theory provides all the correct answers in matters of practical ethics where our intuitions are clear. In difficult cases, the theory either shows that these problems are irresolvable or provides the right answers (the reader is invited to assume that “the right answers” on difficult issues are the ones that she or he believes to be right).

---

<sup>47</sup> Street 2008, p. 220.

Further, the theory correctly validates our beliefs on the relative degree of badness or goodness of various consequences and actions. It turns out that our pre-theoretical judgments of the goodness or badness of acts corresponded to the respective pleasure or pain caused by those acts. More pleasurable consequences are judged to possess more moral goodness than less pleasurable consequences and so on. In addition, the theory helps us see that the cases in which we had previously believed that people were making mistakes usually happen to be cases in which those people misunderstood the tendency of those acts to cause either pleasure or pain. When people come to see that their early judgments concerning the amount of pleasure or pain caused were mistaken, they will change their judgments about the rightness or wrongness of the act. It turns out that our judgments concerning moral goodness (badness) were actually tracking pleasure (pain) even though we were not aware of this.<sup>48</sup>

Let us even pretend that a plausible explanation has been given for why people were unaware of the fact that their moral judgments were tracking hedonic properties. In the distant past there was a conflict between the peoples who are our cultural descendants and those of another tribe whose culture eventually died out. The rival culture apparently advocated living a hedonistic lifestyle, encouraging people to think only of their own pleasure and to ignore the long-term consequences of any of their behavior. (Whether the rival culture really advocated this kind of lifestyle or was misunderstood by our cultural ancestors does not matter.) As a result of this history, there had been a bias against any attempt to identify moral properties with hedonic ones and it took a particularly persuasive theorist to make philosophers take the theory seriously.

Let us assume that on every other moral question people can think of, except normativity, the theory either vindicates our previously held intuitions or provides a convincing explanation for why we might have been mistaken. However, the theory cannot provide an

---

<sup>48</sup> Not all hedonistic moral theories are tracking theories. For example, utilitarianism is sometimes taken to be a revisionary rather than a vindicatory theory.



answer to the question, *Why be moral?*, that will successfully trump all competing non-moral reasons in the way that people like Street demand. In this hypothetical case, I think that we ought to accept that the naturalist has provided a convincing case. The theory stipulates that our moral judgments have been tracking pleasure and pain. Once the initial prejudice against seeing moral properties as hedonic properties dissipates, we will still possess our moral intuitions that killing innocent people is morally wrong and that helping those in need is morally right. Why would we reject the conclusion that killing innocent people is morally wrong just because we see that the wrongness of killing innocent people stems from the pain it creates? Rather what would seem to happen is that we would develop a deeper appreciation for why killing innocent people is wrong.

One reaction we might expect from those who remain skeptical but cannot muster plausible counterexamples would be to maintain that there must be some additional property that was still going unnoticed; a property that causes an act to be good *and* to have pleasurable consequences. Suppose that the property was a natural property. In these cases, the passing of time would work in the theory's favor because one would expect such an additional natural property to be discovered. The more time that passed without the discovery of such a property the less skeptical opponents would become. Some theorists may refuse to accept such a theory, maintaining that goodness must be a non-natural property. It is unclear that such people could be swayed to moral naturalism even if it could provide an account of the overriding and authoritative properties morality appears to have. It is not clear what the moral naturalist could do to convince such a skeptic, nor is it clear that there is any need to.

### Section 1.6 Conclusion

Throughout this chapter I have written of the need to explain moral normativity as one of explaining the connection between moral facts or beliefs and motivations or reasons.

Advocates of Humean psychology naturally tend to equate motivations with desires since, on

the standard Humean account, desires are the only kind of mental state that have the power to motivate. Yet it is not uncommon in the literature to find people speaking of motivations stemming from reasons. One might say that the fact that *X* is morally wrong *provides us with* a reason to refrain from doing *X*, where the type of reason it provides us with is assumed to have motivational force. In subsequent chapters we will try to get a clearer understanding of what kinds of states reasons might be and whether it makes sense to say that they have motivational force. In the next chapter I will look at Michael Smith's work. Reasons play a significant role in his solution to the moral problem.

## Chapter Two

### Section 2.1 Introduction

In this chapter I examine the argument against externalism. I focus on Michael Smith's critique as it has provided a significant challenge for externalists. Central to this challenge is Smith's claim that the externalist's account of the connection between moral judgments and motivation in the morally good agent amounts to a kind of moral fetishism. In the following section I lay out Smith's case against moral externalism. In section 2.3-2.5 I critique three central tenets of the Smith's case: his comparison of color and moral judgments, his political example intended to show the close connection between values and judgments and the claim that *de dicto* moral motivation must be fetishistic. In section 2.6 I discuss alternative accounts of externalist motivation. Section 2.7 includes a summing up of what has been accomplished and looks forward to the work to be done in subsequent chapters.

### Section 2.2 The case for moral externalism

In chapter three of *The Moral Problem* Smith sets out to defend the following claim:

- (1) If an agent judges that it is right for her to  $\phi$  in circumstances C, then either she is motivated to  $\phi$  in C or she is practically irrational.<sup>49</sup>

Smith calls this claim the practicality requirement on moral judgment or 'the practicality requirement' for short. By adding the phrase '*or she is practically irrational*' in (1) Smith wisely avoids some of the problems discussed in chapter one of this dissertation. It allows him to acknowledge that there will be cases like those pointed out by Stocker in which persons will fail to be moved by their judgments of right and wrong. And of course, if judgments of right and wrong do not always motivate they cannot always be actually overriding. We can contrast (1) with (2):

---

<sup>49</sup> Section 3.1, pp. 61-62.

- (2) It is possible that an agent judges that it is right for her to  $\phi$  in circumstances C, be practically rational and lack any motivation to C.

Two readings of (2) suggest themselves. First, we may read (2) as a claim that practical rationality and motivation are *a/ways* contingently linked. Call this global motivational externalism. This seems to be the common understanding of externalism. For example, Brink writes that “externalism claims that the motivational force and rationality of moral considerations depend upon factors which are external to the moral considerations themselves.”<sup>50</sup>

Externalists often appear to view the mind/brain as a place in which there are moral judgments that are produced in one area and, added to this is a general desire to do that which is right which is located in another area. If this is correct then for each moral judgment that is made it will be a contingent matter whether it is accompanied by motivation. Thus, global motivational externalism will be true. However, (1) and (2) do not exhaust the possibilities in this area, for it may be that some moral judgments or concepts necessarily carry motivation with them whereas others do not. Consider (3), which I will call local motivational externalism:

- (3) In some (but not all) situations it is possible that an agent judges that it is right for her to  $\phi$  in circumstances C, be practically rational and lack any motivation to  $\phi$ .

How it is possible for local motivational externalism to be the case will be explored in subsequent chapters. I merely point it out here to show that such a view is logically possible. What a defense of (3) will require is a naturalistic perspective on moral motivation. To anticipate, it may be that moral information can be processed by the brain in different ways. Perhaps some ways of processing moral information lead to a physical connection within a particular moral agent between the fact that X is morally right (wrong) and the corresponding motivation such that the one cannot occur without the other whereas other ways of processing this information leads to a contingent connection between the acquisition of a moral belief and motivation to act on that belief. This possibility calls to mind the Aristotelian distinction between

---

<sup>50</sup> Brink 1986, p. 28. See also Brink 1989, p. 42.

morally continent or incontinent agents and those that are fully virtuous. The possibility of this type of solution will be explored in the fourth and fifth chapter of the dissertation.

Now I want to return to Smith's defense of (1). Smith spends a considerable amount of time addressing Brink's amoralist objection. The task is important to Smith's project for two reasons. First, Smith wants to maintain that sincere moral judgments are necessarily accompanied by motivation. If this is correct then there cannot be any people who are amoralists. An amoralist is someone who makes a moral judgment but remains totally unmoved by that judgment. If amoralists exist, then (1) is false. Second, Smith wants to maintain that it is an a priori truth that moral judgments motivate. This requires even more of him. He must maintain that the idea of an amoralist is somehow incoherent and that amoralists are not even conceivable.

Those who object to the possibility of the amoralist often do so on the grounds that there may be people that we think of as being amoral but that they fail to make moral judgments. When they judge that an action is right or wrong they most likely are making judgments about what others call right or wrong. This amounts to a-less-than-sincere judgment about rightness or wrongness.<sup>51</sup> Smith agrees with Brink that Hare's strategy as it stands will not work and that the amoralist seems conceivable.<sup>52</sup> Hare argues that people who make inverted commas judgments of right and wrong are making judgments about how others use the words. But there does not seem to be any conceptual reason why amoralists cannot use the words to talk about what is really right; we do not need to see them as changing the subject, so to speak. Smith suggests that we will be better off seeing amoralists as people who "try to make moral judgments but fail."<sup>53</sup> That is, we should not try to identify a different type of judgment that the amoralist is actually making.

---

<sup>51</sup> See Hare 1952.

<sup>52</sup> See Svavarsdottir 1999, pp. 176-180 and pp. 188-189 for a well-drawn example of an amoralist who apparently makes moral judgments.

<sup>53</sup> Smith, section 3.4, p. 68.

To try to spell out more clearly the difficulty concerning whether we ought to believe that the amoralist is really making moral judgments, Smith develops an analogy with color. The analogy is important because it brings to light the theory dependence of the intuitions on both sides of the debate. Imagine that someone who has been blind from birth can reliably use color terms because he has been hooked up to a machine that allows him, through his sense of touch, to identify the colors of different objects. Does such a person possess the concept of (e.g.) red? Those who wish to give an affirmative answer can point to the person's reliable use of 'red' to identify red objects. Others will object that because the person is lacking a certain kind of visual experience, roughly speaking, the experience of 'redness' the person does not possess the concept. Both sides could rightly be accused by the other of begging the question. One side assumes that reliability is sufficient for concept mastery whereas the other assumes that the appropriate visual experience is necessary to have the concept 'red'.

Similarly, the externalist wants to maintain that the moral judgments made by the so-called amoralist really are moral judgments. According to the externalist, the fact that they are not accompanied by the motivation we are accustomed to is no reason to doubt their genuineness. If the amoralist can reliably identify right and wrong then she possesses the concept. On the other hand, the internalist believes that the amoralist must not be making genuine moral judgments because sincere moral judgments are accompanied, absent weakness of the will, etc., by motivation to act in accordance with those judgments. There is a certain sort of experience that the amoralist lacks that the internalist believes is necessary for moral judgment. But what independent reason does the internalist have for this conviction other than her ideological belief?

Smith argues that only the internalist can give a convincing explanation of motivation in the morally good person. Smith starts by noting that "it is a striking fact about moral motivation that a *change in motivation* follows reliably in the wake of a *change in moral judgment*, at least

in the good and strong-willed person”<sup>54</sup> and illustrates this with a hypothetical example in which he is involved in a political debate with a friend. Smith enters the debate with a conviction that he ought to vote for the libertarian party whereas his friend believes that he ought to vote for the social democrats. Smith’s friend convinces him that he ought to vote for the social democrats “not just because the social democrats will better promote the values that (Smith) thought would be promoted by the libertarians, but rather because the values (Smith) thought should and would be promoted by libertarians are themselves fundamentally mistaken.”<sup>55</sup> *In the wake of* (this is Smith’s phrase) these new value judgments, new motivations will follow.

Who is best situated to explain this change in moral motivation? Internalists can say that the moral judgment itself is capable of providing the motivation. When Smith judged that voting for the libertarians was right he was motivated to vote for the libertarians. When his judgment changed, his motivation changed accordingly. The externalist seems to be required to give a different story of what happened. According to externalists, moral judgments in and of themselves do not have the power to motivate. The motivation must come from somewhere else. A reasonable explanation of the change in motivation for the good and strong-willed person from the externalist’s point of view is that this person has a standing disposition to do *what is right, whatever that happens to be*.

To understand the distinction between externalist and internalist accounts of moral motivation it will be helpful to look at a distinction Smith introduces between de dicto and de re motivation. As the terms ‘de dicto’ and ‘de re’ are being used outside of their typical context, some explanation of Smith’s use is warranted. These terms most often occur in discussions of the scope of the modal property in question. Sometimes the scope of the property ranges over the entire sentence, as in (1) below and sometimes it ranges over a part of the sentence, as in (2) below. In the first case it is called de dicto modality and in the second de re modality.

---

<sup>54</sup> Section 3.5, p. 71.

<sup>55</sup> Ibid.

- (1) Necessarily, the smartest person is smarter than any other person.
- (2) The smartest person is necessarily smarter than any other person.

(1) is true but (2) is false because the necessity in (2) attaches to the particular person. The smartest person, whoever he or she is, could have turned out to not be the smartest person. For example, another person could exist who was smarter than the person who actually is the smartest. Put another way, there are possible worlds on which the person who happens to be the smartest person in the actual is not the smartest there but, whoever the smartest person is on any possible world has to be smarter than any other person on that world.

However, there are true modal de re claims, such as the following:

- (3) The number of days in a (non-leap) year is necessarily 365.

In (3), the necessity attaches to the number of days in a (non-leap) year and says that this number of days is necessarily 365 but (3) is false on a de dicto reading because the number of days in a (non-leap) year could have been more or less than 365. The Earth might have had a different orbit or rotated on its axis more or less quickly.

Smith uses the de dicto/de re distinction to refer to the scope of motivation rather than the scope of a modality, such as necessity or possibility. When he speaks of moral motivation as being de dicto, he means that if motivation is this way for a person then the person has a *motivation to do what is right, whatever that is*.<sup>56</sup> In a way somewhat similar to the way necessity can apply to an entire proposition, Smith is indicating that a person would, if externalism is true, attach their motivation to a particular proposition. If each of us possesses de dicto motivation then each of us is walking around, so to speak, with a desire to do what is right but the desire is not attached to any particular judgment regarding some particular course of action. Now suppose that I hear about a rally for equal rights taking place at the capitol building this afternoon. I judge that it would be right for me to attend the rally. This judgment by

---

<sup>56</sup> I am borrowing this phrase from Sadler 2003.



itself will not motivate me to attend the rally. What will motivate me is my separate desire to do what is right whatever that turns out to be.

On the other hand, *de re* motivation is attached to particular moral judgments. If I am motivated *de re* then when I judge that it would be right for me to attend the rally this judgment itself is sufficient to motivate me to attend the rally. It makes sense for Smith to defend this type of strategy given how he construes the connection between moral beliefs and motivation. Smith sees a judgment about what it is rational for a person to do in a situation as being what gives rise to the motivation. So, if I judge that it is right to attend a rally for equal rights then, assuming I am not irrational, I will be motivated to attend the rally. The specific judgment about the rightness of supporting equal rights creates the motivation. Hence, Smith's account favors *de re* motivation.

If the externalist is right then when we are motivated to do a particular action that we believe is right this motivation is derived from the standing desire to do what is right, whatever that may happen to be, coupled with our judgment in a particular case. Smith then makes the bold claim that the strong externalist's position leads to a straightforward *reductio*.<sup>57</sup> This is so, Smith tells us, because "good people care non-derivatively about honesty, the weal and woe of their children and friends and fellows", etc.<sup>58</sup> Phenomenologically speaking, Smith has a point. It certainly seems strange if, upon judging that my best friend is in danger, I am moved to help her not directly from that judgment but in conjunction with an additional judgment that it would be morally right to do so. Until it occurs to me that it would be morally right to help my friend I feel no motivation to help her. Smith claims that to be motivated in such a way "is a fetish or moral vice"<sup>59</sup> and notes its similarity to Bernard Williams's objection to the claim that the moral person must be motivated by impartial concern because it gives the person *one thought too*

---

<sup>57</sup> Section 3.5, p. 75.

<sup>58</sup> *Ibid.*

<sup>59</sup> *Ibid.*

*many*. The strong externalist requires that we possess an impartial concern to do the right thing whatever that may be in addition to our particular moral judgments.

I think that Smith's claim that moral motivation must be *de re* rather than *de dicto* in order to be plausible is incorrect for a number of reasons. In what follows I will examine three crucial aspects of Smith's critique of externalism: the analogy with color, the political example he offers, and his claim that *de dicto* motivation is fetishistic. I will argue that there are serious problems with these three aspects of his argument and that this undermines his argument against externalism.

### Section 2.3 Critiquing the color analogy<sup>60</sup>

First, Smith claims that an independent reason is needed to decide whether we should accept whether amoralists truly make moral judgments just as we would need an independent reason to decide who is right about color judgments. One way that we might try to decide is to draw upon a more comprehensive theory. For example, we could appeal to an epistemological theory that explains concept possession by appealing to the agent's ability to successfully reidentify the individual, stuff or natural kind to which the concept applies.<sup>61</sup> The agent acquires the concept by being able to track the individual, stuff or natural kind.<sup>62</sup> Since moral naturalists who are realists tend to identify goodness with one or more natural kinds, this sort of view may be attractive to them. Having a concept of red or goodness would then be a matter of being able to reliably reidentify that kind. This would make reidentifying goodness or redness the same type of process as identifying other substances. David Copp provides an account of moral realism of this variety which he calls a quasi-tracking account.<sup>63</sup> So the moral realist who

---

<sup>60</sup> See Brink 1997, pp. 22-26 for additional criticisms of Smith's use of the color analogy.

<sup>61</sup> See Millikan 2000, ch. 1 for an introduction to this type of approach.

<sup>62</sup> See Millikan 2000 but also see Millikan 2004, especially ch. 3 and ch. 4.

<sup>63</sup> See Copp 2008. I do not mean to imply that Copp agrees with Millikan's theory.

defends externalism can appeal to a more general epistemological theory to defend his theory. Of course, there is nothing to stop the internalist from employing a similar strategy.

I admit to being somewhat confused by Smith's color analogy. First, it might be thought to bias the reader in favor of an internalist account because the person who supposedly knows the good without being motivated by it (i.e. the amoralist) is supposed to be like a person who knows what red is without ever having seen red. It does seem somewhat counterintuitive to say that someone knows what red is without ever having seen red. In addition, just as a seeing of red is typically (always?) accompanied by a phenomenological experience of redness, a judgment that X is morally good is typically accompanied by a feeling of motivation to do X.<sup>64</sup> However, it is not clear that the motivation is somehow a part of the concept. What seems to be going on is that we are making a judgment and the motivation is in some way the effect of the judgment. It may be helpful to distinguish between two ways that moral facts and moral motivation may be connected.

1. If X makes a moral judgment that A is good then X will be motivated to do A.
2. If X makes a moral judgment that A is good then X will judge that people ought to do A.

(1) seems to be true for the most part in that there appears to be a reliable connection between moral judgments and motivation. It does seem that when we make moral judgments there is some motivation to act in accordance with them, even if that motivation is overridden by other motivations we have. Notice that it is when we suffer from various maladies such as depression or addiction or just plain tiredness that our moral judgments do not seem to be accompanied by their typical motivating force. Yet, if (1) is the correct way to think of the connection between moral facts and moral motivation then it does not seem that being motivated is part of our *concept* of the good. To see this, consider the following. Suppose that we were going to define our concept of the good in terms of the platitudes that typically accompany goodness. If we

---

<sup>64</sup> There is a phenomenon called blindsight in which people can respond to visual stimuli without having conscious awareness of the stimuli. I will not explore the relevance of this issue here.

were gathering up our platitudes about the good, (1a) seems a more plausible candidate than (1b).

1a. If an agent realizes that X is good then that agent will *ceteris paribus* be motivated to do X.

1b. If an agent realizes that X is good then that agent will *necessarily* be motivated to do X.

After all, weakness of the will has been an issue dating back at least to Aristotle. However, suppose that the internalist insists that (1b) is preferable to (1a). The internalist seems to be wrong for at least two reasons. First, he is ignoring the empirical facts. The good does not always motivate. If we are gathering our platitudes in an attempt to define the good, (1b) will only be acceptable if we are ignoring empirical evidence. What would be the justification for ignoring empirical evidence?

Second, both (1a) and (1b) are about the effect the good has on those who grasp it. Yet, it seems implausible in most cases to think of the effects of something as a part of the concept of that thing. Cancer is a common effect of (heavy) smoking yet it is strange to think of cancer as being a part of the concept of smoking. Pollution may be a common effect of industry but we do not think of pollution as being a part of the concept of industry. It may be that these become linked in our minds for various reasons, perhaps because they co-occur so frequently or because certain effects are particularly salient. However, they typically remain separable. Not all smokers get cancer, we can think of industries as clean, and the good does not always motivate.

It is true that in certain cases we classify something by the effect that it has. For example, our concept of a poison is of something that causes sickness or death. Yet we commonly think of poisons under other concepts as well. Engine coolant is a poison and it is not uncommon for pets to die from drinking it. However, it would be strange to think of being poisonous as being a part of the concept of being engine coolant. Consider. If the composition

of engine coolant were changed so that it was no longer poisonous we would not be inclined to think that our concept of engine coolant had changed.

Perhaps we would want to say that morality is different because the moral judgment and moral motivation are so closely linked in our minds that we just cannot separate them. To make this claim at all plausible, it would seem that the internalist must at least hold that the two are always linked and that (1b) more accurately captures our experience with morality than (1a). Yet the admission that various maladies can affect moral motivation amounts to an admission that the motivation is not a necessary aspect of our concept of moral judgment. For these reasons, I do not think that (1) captures the link between motivation and goodness that is arguably inherent in our concept of goodness.

On the other hand, (2) seems to do a better job of capturing what would be a necessary condition for possessing the concept of goodness as it makes a connection between moral judgment and judgments about behavior. Suppose we hear about a person who says that he knows that torturing and killing people is morally wrong but sincerely maintains that he does not feel any motivation to refrain from doing so. We might ask if he understands that people ought to do as they are morally required and so on. He may very well understand this and yet not feel any motivation to follow certain moral rules.

If (2) is a more likely candidate for a necessary condition of moral goodness than (1) then the color analogy breaks down. If having a phenomenological experience of (e.g.) 'redness' is a necessary condition for possessing the concept 'red' then the necessary conditions for having a concept of 'goodness' differ from the necessary conditions for having the concept of 'redness' in a way that is detrimental to the analogy. On the other hand, if (1) better captures the necessary conditions for having a concept of goodness then we are forced into a confusing situation. What are we to say when a depressed person concludes that she ought to do X but feels no motivation to do X. Has she lost a concept that she formerly possessed? This

seems highly implausible. Based on the preceding considerations, I conclude that Smith's analogy is faulty.

#### Section 2.4 Critiquing the political example

I think that it is right to say that moral motivation frequently accompanies a change in moral judgment but that the example which Smith gives to illustrate this is implausible. This will become clear if we spell out the example in some detail in order to critically examine it. I spend significant time on this example for two reasons. First, it has largely been neglected in the literature on Smith.<sup>65</sup> Second, it helps to illustrate a distinction between valuing and decision making. I will modify it somewhat to address American political parties as I am more familiar with the values that they promote.<sup>66</sup>

Suppose I am having a debate with a friend about which political party we ought to vote for in the next election. I am convinced that people ought to vote for the Libertarian Party whereas my friend is convinced that people ought to vote for the Democratic Party.<sup>67</sup> I come into the discussion believing that one ought to vote Libertarian because they promote the values of self-reliance, hard work and freedom. I think my friend is mistaken because he thinks that the government ought to do more for the people, which makes them less self-reliant and requires them to work less hard. Further, if his party is able to pass the legislation that they wish to pass, taxes will be raised and various new regulations will be established, both of which I see as limiting people's freedom to live their lives as they see fit. When it comes to specific policy issues, I support the libertarians because they are opposed to government involvement in health care and believe the educational system should be privatized.

---

<sup>65</sup> However, there is a helpful discussion of the example in Shafer-Landau 1998, who uses the example to show that motivation can be external without being de dicto. This will be discussed in section 2.6 below.

<sup>66</sup> Nothing philosophically crucial hangs on altering the example in this way.

<sup>67</sup> The libertarian candidate is rarely a viable option in elections in the United States of America. For sake of example, let us pretend that the candidate routinely is viable. Again, nothing crucial rests on this assumption.

There seems to be two basic ways that the conversation between me and my friend might go. The first is to make me see that my fundamental values are compatible with different policy positions.<sup>68</sup> Suppose that my friend convinces me that, for many people, no matter how hard they work they cannot afford to provide health insurance for themselves and their families. Providing government sponsored health care will not discourage people from having a good work ethic. Also, my friend tells me that, while he generally favors self-reliance, he does not think it is relevant here. He agrees that people should strive for self-reliance in the areas in which it is reasonable to be self-reliant but that there are some areas in which we cannot really be self-reliant. He tells me that education is this way for many of us. If we did not learn physics or algebra in school, how can we be expected to teach it to our children? Though some of us will be able to afford to send our children to private school, without public school many of our children will be inadequately educated. Finally, my friend suggests that while he too values personal freedom, the kind of intrusion and limiting of freedom that is involved with public education and guaranteed health insurance is relatively minor and is worth the benefit of enhanced public health, having an educated populace and the peace of mind that comes from knowing that we have these safeguards. My friend's argument has convinced me and I now intend to vote democrat. In this case, I do not really change my fundamental values. I still value self-reliance, hard work and personal freedom but now I recognize that I have been drawing upon them in situations in which they are not relevant. My motivation was to promote these values and I thought that doing my part to elect a libertarian candidate was a way to accomplish this. My desire to vote libertarian was instrumental to my desire to promote my core values. Now I see that I was wrong. Voting libertarian will not do more to promote my core values—or not much more at any rate—than voting democrat would. This coupled with other considerations leads me to become motivated to vote democrat.

---

<sup>68</sup> I will say more below about what it means to be a fundamental value. Smith uses this term but does not elaborate on exactly what a fundamental value is.

Though this seems a plausible way that my friend could change my stance, this does not seem to be the way Smith intends the example to be read. Smith is clear that this is to be a conversation in which one's *most fundamental values* are changed. So a second way that my friend might convince me to vote for the Democratic Party is to convince me that my values themselves are wrong. Self-reliance, hard work, and personal freedom are not values that I ought to have. This seems to be the reading that Smith has in mind for his example. And each of these values is open to debate. There are different societies in which these values are not widely embraced. It is not irrational in any obvious way to value interdependence over self-reliance or security over freedom. Though hard work may sometimes be wise because it helps one get ahead, it is not clear that placing more emphasis on leisure and spending time with one's family cannot also be important values, values which may take precedence over hard work.

If this is the way Smith intends us to view his example then it becomes problematic because it is either implausible or it does not show what he intends it to show. It will be useful to discuss the nature of valuing to demonstrate why this is so. Smith (1994) includes an extended discussion of value.<sup>69</sup> Before addressing Smith's discussion, let me make a few intuitively plausible comments concerning the nature of valuing. First, when we say of A that she values X, there is an assumption that A believes that X is valuable or good. This is because there are some things that we desire that we do not value. The heroin addict may desire more strongly than anything else a fix and yet hate himself for giving into the desire. He believes that being sober is of value even though he struggles to maintain his sobriety. Yet there seems to be a desiring aspect to valuing as well. If the heroin addict values sobriety then it is reasonable to assume that he feels some sort of pro-attitude toward sobriety, even if this pro-attitude is frequently undermined by an intense desire for a fix. Consider the following necessary condition for a value.

---

<sup>69</sup> See sections 5.3-5.8.



( $\Psi$ )( $\mathcal{E}$ ) If  $\Psi$  is a value and  $\mathcal{E}$  is a valuer, then if  $\mathcal{E}$  possesses  $\Psi$  then  $\mathcal{E}$  believes that  $\Psi$  is valuable and  $\mathcal{E}$  has a pro-attitude toward  $\Psi$ .

I do not want to suggest that the condition above is also sufficient and I will not try to establish sufficient conditions for value. Samuel Scheffler argues that in addition to the above, a value also includes a tendency to experience a certain range of emotions, among other things.<sup>70</sup>

The necessary condition above gives us prima facie reason to think that a value contains elements of both beliefs and desires. J. E. J. Altham suggests that moral mental states might be thought of as states which have both directions of fit.<sup>71</sup> These “besires” are like beliefs in that they tell us how the world is and like desires in that they direct us to change the world in a certain way. Smith argues that there can be no such thing as besires but his arguments are unpersuasive. To anticipate, I will argue in chapter three that Smith’s case against besires fails because it does not consider the possibility that besires may exist alongside distinct moral beliefs and moral desires. In chapters four and five I argue in favor of a dual-state view which includes besires as well as moral beliefs and moral desires.

The argument against besires plays a key role in Smith’s overall project and his failure to prove that they do not exist undermines his project. This is because his argument leads him to dismiss the possibility that valuing could be something like a besire. After rejecting attempts by Donald Davidson, David Lewis, and David Gauthier to explain valuing in terms of desires, he says that “if we cannot reduce valuing to desiring then we have *no alternative* but to consider reducing valuing to believing.”<sup>72</sup> Thinking of valuing in terms of believing leads Smith to the conclusion that to have a value is to have a belief about what we would have a desire to do if we were fully rational. These beliefs then lead their possessors to have desires to act accordingly, assuming their possessor is rational. Smith’s view looks more plausible if we eliminate the

---

<sup>70</sup> See Scheffler 2011.

<sup>71</sup> See Altham 1986.

<sup>72</sup> Smith 1995, p. 147. The italics are mine.

possibility that valuing is a state with aspects of both beliefs and desires, as valuing initially looks to be.

Smith refers to fundamental values but does not elaborate on what makes a value fundamental. Qualifying a value as fundamental makes it reasonable that Smith intends to be distinguishing this kind of value from others. I here make my own distinction between two kinds of values which I will call shallowly-held values and deeply-held values.

Df: If X should be judged to be a shallowly-held value for an agent A then A claims to value X but shows little or no interest in promoting X, especially if doing so will come at some significant cost (monetarily or otherwise) to A.

Df: If X should be judged to be a deeply-held value for an agent A then A shows an interest in promoting X even when doing so comes at a significant cost (monetarily or otherwise) to A.

These definitions have been formulated to show what might plausibly be considered necessary conditions for the judgment that an agent possesses a shallowly-held or deeply-held value. Many of us have values that we hold shallowly. If asked whether we care about curing some disease or helping people in a faraway land many of us will say that we do and affirm that it would be good to cure the disease or help the foreigners but exhibit little if any motivation to actually do so. On the other hand, some people do behave in ways that demonstrate that they care about these kinds of issues and that they will commit a great deal of time and/or money to them. Now let us return to my reformulation of Smith's political example.

If my deeply-held values do not change when I decide to vote for the democratic candidate instead of the libertarian candidate the example is plausible, as is the accompanying change of motivation, but the example does not show what Smith wants it to. On this interpretation, the reason I was motivated to vote for the libertarian candidate was because I saw doing so as a means to an end of promoting my deeply-held values of self-reliance, hard work and freedom. My desire to vote libertarian was merely instrumental. As soon as I realized that voting libertarian will do little to promote these values and that I have other good reasons to vote democrat, my motivation changed. In this case the moral motivation is more plausibly

thought of as attaching to the deeply-held values and not the judgment about which party's candidate I will vote for.

Yet Smith is careful to stipulate that the political conversation causes a change in one's most fundamental values. It seems plausible to assume that a fundamental value is like a deeply-held value. Now were my friend to actually cause such a shift in my fundamental values then it seems likely that a motivation shift would follow. Further, it would not be unnatural to characterize such motivation as *de re* rather than *de dicto*. Unfortunately for Smith, the example becomes deeply implausible when we are forced to understand it as a shift in our fundamental values. This becomes acutely obvious in cases in which we recognize that we have a value that we would rather not have. Shannon Sullivan describes the following example.<sup>73</sup> While attending a conference in Alabama, she met a white man (Charlie) at a bus stop. (Sullivan is also white.) The man boarded the bus with her and sat in the seat behind her. When a number of black people boarded the bus, the man took the seat next to her so that she would not have to sit next to "someone she did not know" (i.e. a black person). The man's behavior presumably has its roots in the idea that it is necessary to protect white women from black men. However, as Sullivan points out, it is not clear that the man would agree with this characterization of his behavior and he might be deeply offended by it.

Let us embellish the story and assume that the man was raised by explicitly racist parents. As a result, he came to have numerous deeply-held values which were also racist. Among these deeply-held values was the value of protecting white women from a threat he believed black men posed to them. As a college student, however, Charlie had a deeply moving conversation with an African-American student who challenged his racist values. From that day on, he has reported that people ought to be treated equally, regardless of the color of their skin and becomes angry if someone suggests otherwise. Yet he still sometimes behaves in ways that are reminiscent of his racist past, even if he is unaware of doing so. The

---

<sup>73</sup> Sullivan 2006, pp. 114-115. I am indebted to Laura Papish for bringing this example to my attention.

embellished story does not strike us as implausible. In fact, it is far too common.<sup>74</sup> Deeply-held values do not change easily, as many have discovered when they find themselves having values of which they do not approve.

Deeply-held values are problematic for Smith because a change in moral judgment is supposed to bring with it a change in moral motivation. Interestingly, a dual-state theory that embraces both externalist and internalist types of motivation seems to afford a more natural explanation of what is going on here.<sup>75</sup> Shallowly-held values, which we may think of as being formed anytime we make a judgment that we ought to do something, are accompanied by either relatively weak motivation or none at all. (Such judgments need not always be instrumental means to promoting our deeply-held values.) Understandably, a change in moral judgment in such cases generally brings about a change in motivation because there is little or no motivation accompanying the original judgment. This phenomenon may be explained by a standard theory of beliefs and desires. On the other hand, we may think of deeply-held values as being a type of besire. In such cases, the difficulty in changing motivation when a different moral judgment is made may be explicable by citing two different moral systems at work in which there is a conflict between a moral belief and a besire. I will explore these themes in more detail in chapters four and five.

### Section 2.5 Critiquing the fetishism argument

In this section I will argue that de dicto moral motivation is not fetishistic. A number of people have argued that Smith is mistaken about the strangeness of de dicto moral motivation.<sup>76</sup> I will discuss three cases here. First, there are cases in which an agent is uncertain about what the right thing to do is. In such cases, an agent cannot be motivated by de

---

<sup>74</sup> Of course, it is better to renounce racist values and strive to be non-racist and fall short than to never renounce the values and make the effort in the first place.

<sup>75</sup> I will not explore this in-depth here.

<sup>76</sup> For example, see Lillehammer 1997, Olson 2002, Sadler 2003.

re considerations because she does not know what is the right thing. However, suppose she also lacks a *de dicto* motivation. In such a case, she will lack any motivation to figure out what the right thing to do is. Yet it is not strange to think of morally good agents as being ones who have a standing concern to do what is right that can motivate them to figure out which actions are right. Such a person may recognize that there are morally relevant factors present but these factors point to different and conflicting responses. Possessing *de dicto* motivation could lead the agent to engage in moral deliberation.<sup>77</sup> It could also motivate the agent to monitor the situation for additional information that would help her to determine what the right thing to do is.

A second situation in which possessing *de dicto* motivation is not fetishistic involves situations in which we are tempted to do the wrong thing.<sup>78</sup> The following example is from Lillehammer 1997.

Consider someone who goes to a party during a phase when she is tired of her husband. At the party she meets a very charming person and is tempted to have an affair. She judges that it would be wrong to have an affair on account of her husband's feelings. But she is temporarily indifferent to her husband's feelings. However, she has a standing *de dicto* desire to do what is right which, together with her moral judgment, causes her to do the right thing, in spite of the absence of a *de re* desire to do the right thing and the presence of a *de re* desire to do the wrong thing.

The problem for the woman in the example is that she has conflicting motivations and her *de re* concern for her husband is either anemic or completely absent. In such a case, a desire to do the right thing where this is read *de dicto* instead of *de re* can help a person to do the right thing. Nor would we judge the woman's motivation to be fetishistic in this case. Her desire to stay faithful to her husband in spite of her temptation to do otherwise seems to be reason for admiration rather than condemnation.

A third issue that arises is the need in the good person for *de dicto* motivation in order for her to be able to periodically examine her *de re* motivations.<sup>79</sup> Call this the immoral *de re*

---

<sup>77</sup> Sadler 2003, p. 69-70.

<sup>78</sup> For examples see Lillehammer 1997, p. 192, Olson 2002, pp. 93-94, Sadler 2003, pp. 70-71, and Toppinen 2006.

motivation objection. This objection is related to the previous one in that there is a conflict between a *de re* motivation and what the right thing to do is but it differs in that it involves being motivated in a way that we normally judge to be good. For example, to be moved by a direct concern for a loved one may in general be a good thing but examination of these motives is also morally necessary. Suppose I need to hire someone for a job and I know that my brother would like to have that job. I feel motivated to hire my brother by a direct concern that I have for him. However, if I am a good person I should ask myself if it is morally permissible for me to hire my brother in this situation. Suppose I decide that it is morally permissible and I hire my brother. I tell him that I did so because, having given it some thought, I decided to hire him because he is my brother *and* it was morally permissible for me to be partial in this case. If my brother is offended by my having thought about what was morally permissible in this case, it arguably indicates a moral failing in my brother rather than my having *one thought too many*.<sup>80</sup> It would be strange if the fact that something would positively impact my brother's welfare did not motivate me but it is not morally repugnant to consider whether acting on that motivation is morally acceptable in this situation. Taken to an extreme, such direct concern for family and friends can lead to quite immoral behavior. Suppose I come to have very good reason to believe that my brother is a serial killer. Yet I do not turn him in—after all, I didn't know any of his victims so what are they to me?

This leads us to Smith's objection that *de dicto* moral motivation is odd because it requires us to have "one thought too many".<sup>81</sup> Bernard Williams used this phrase in an objection to utilitarianism. There are times when utilitarianism would require us to ask a question or have a thought that it normally seems that we ought not to have. This occurs in personal relationships in which it is plausible that it is morally acceptable to show favoritism. If followed closely, an impartial theory like utilitarianism might require one to ask, "Is my helping

---

<sup>79</sup> Brink 1997, pp. 27-29 also discusses this issue.

<sup>80</sup> This phrase is from Williams 1976 and will be discussed in more detail below.

<sup>81</sup> See Smith, section 3.5, p. 75.

out my loved one right now the best thing I can do to maximize utility?" Such a question seems inappropriate and our loved ones may understandably be offended by our having such a thought.<sup>82</sup> It may seem to the loved one that we ought to be directly moved by the concern for that loved one without intermediate impersonal concerns about moral rightness. Arguably, we do not need to think about these things in many situations.

Smith refers to Williams' objection as one that is related to his own concerning de re and de dicto desires. According to Smith, the moral externalist winds up "elevat(ing) a moral fetish into the one and only moral virtue."<sup>83</sup> But Williams' objection is concerned with deliberation.<sup>84</sup> In the case in which a utilitarian moral agent has one thought too many, this thought would normally be accessible to introspection. Not only that, the thought would generally be one that we were aware of having as we were having it. Perhaps an experienced utilitarian thinker could go through the reasoning in such a way that the thought processes would be automatic and not enter his conscious mind but he could also go through the reasoning consciously if need be. Are our processes of moral motivation similarly accessible to introspection?

One way of thinking of motivation suggests an affirmative answer. We are frequently aware of non-moral desires such as a desire to eat something, as well as moral desires such as the desire to give to charity or to help a friend in need. Unfortunately, this account of moral motivation conflicts with the one offered by Smith. To see this it will help to take a closer look at his account of motivation. Recall that Smith distinguishes between two types of reasons, normative reasons and motivating reasons. Motivating reasons are psychological states which can be appealed to in order to make our acts intelligible. Motivating reasons are primarily explanatory as opposed to justificatory. Smith accepts a Humean theory of motivating reasons.

In defending the Humean theory of motivating reasons Smith rejects a phenomenological conception of desires. First he considers what he calls the strong

---

<sup>82</sup> It seems that they should not always be offended though. See my example in the previous paragraph.

<sup>83</sup> Smith, section 3.5, p. 76.

<sup>84</sup> See Joyce 2001, pp. 64-66.

phenomenological conception of desire and defines it as follows: “desires are, like sensations, simply and essentially states that have a certain phenomenological content.”<sup>85</sup> He rejects this account for three reasons. First, it is possible to have a desire and yet not believe that we have that desire. Smith gives the example of someone who goes out of his way to go to a newsstand because there is a mirror there which enables him to look at himself but who cannot admit to himself that this is his motivation for going to that particular newsstand. The person desires to look at his own reflection but does not believe that he desires to do so. The second reason Smith gives for rejecting this account is that we can be confused about the desires that we possess. Smith gives the example of a man who claims that it is “one of his fundamental desires . . . to be a great musician.”<sup>86</sup> If asked, this man would deny that his desire is really a desire to please his mother. However, when the man’s mother dies his desire to be a great musician vanishes. We can safely assume that it was the desire to please his mother that was motivating his behavior, *whether or not the man ever comes to realize it*. The third reason Smith gives for rejecting the strong phenomenological account is that desires differ from sensations in that desires have propositional content whereas sensations do not.

Smith also rejects a weaker phenomenological account in which it is essential that a desire possesses phenomenological content. The earlier examples help to illustrate this but Smith gives one more example. We would be inclined to attribute to a father who goes out of his way to promote his children’s long-term welfare a desire that his children do well in life. The man may on some occasions report that he has such a desire though it would be implausible to say that the desire is something that the man is aware of at all times.

Having rejected a phenomenological account of desire Smith adopts a dispositional account instead. In developing what he calls a Humean view, Smith says that “we can find in Hume’s suggestion about the epistemology appropriate for the calm passions (that they *are*

---

<sup>85</sup> Smith, section 4.5, p. 105.

<sup>86</sup> Smith, section 4.5, p. 106-107.



*more known by their effects than by their immediate sensation*<sup>87</sup>) the inspiration for a somewhat different conception of desires.”<sup>88</sup> More could be said about Smith’s account of desires but I now wish to return to Smith’s distinction between de dicto and de re motivation.

The example of someone helping a friend from impersonal motives rather than a direct concern is most objectionable when we think of motivation in conscious terms. I see my friend in trouble and become consciously aware of being motivated to help her not because I care about her but because there is nothing else I could be doing at the present time which would bring about a greater amount of moral good. After helping her, I reveal my motive for helping her. She is offended. But if Smith is right about the nature of desires, there is no reason to believe that our moral motivations will always be present to consciousness, even if they sometimes are. This leads to an epistemic problem. How are we to determine whether our motivation is de dicto or de re in cases in which we are not consciously aware of our motivations? Call this the introspection problem.

### Section 2.6 Alternative explanations of the close connection between moral judgment and moral motivation

First, it should be noted that if de dicto moral motivation is not fetishistic, and we have seen good reason to believe that it is not, it can provide an acceptable account of moral motivation in the morally good person. What is necessary in accounting for motivation in the morally good person is that we can explain the reliable connection between that person’s making a judgment and being motivated by that judgment. As Smith rightly points out, on the externalist picture our motivation to (e.g.) vote libertarian must be derived from our judgment that it is right to do so. To see this, consider a case in which a judgment is not. Suppose that I am motivated to drink soy milk because I like the taste better than that of animal milk. Later I

<sup>87</sup> Smith, section 4.5 p. 113, quoting Hume.

<sup>88</sup> Smith, section 4.6, p. 113.

come to believe that the consumption of animal milk is morally wrong because of how the animals are treated. Though my moral judgment and motivation match up, they do not do so for the right reason. If I am a morally good person I ought to be motivated to avoid animal milk *because* it would be right to do so. In the example described above, the matching up of judgment and motivation is a happy accident. (Of course, I may still be a morally good person whose moral judgment would have led to a motivation to avoid animal milk if I had not previously become so motivated for reasons of taste.)

The morally good person with de dicto moral motivation will reliably experience a motivational change when she changes her judgment. As was pointed out in the previous section, it is inconsistent of Smith to characterize externalist moral motivation as a “self-consciously moral motive.”<sup>89</sup> We can have a de dicto desire to do what is right whatever that turns out to be yet lack conscious awareness of this much of the time, just as a father may have a long-term desire that his children do well in life without being continuously aware of it. The desire may be one that we become consciously aware of only in certain circumstances, such as when there is confusion about what the right thing to do is or when we have a conflict between our desires.

Though I believe the considerations discussed in this chapter should be sufficient to alleviate our concerns about de dicto motivation, it should be noted that there are options for the externalist who, for one reason or another, rejects the idea of de dicto motivation. For example, James Dreier expresses concern that the fetishism argument is largely effective against de dicto motivation.<sup>90</sup> He develops an alternative model in which the morally good person is someone who possesses a second-order desire to acquire first-order desires to be moved by the right making features of potential actions. Once we have attained a first-order desire it can be this desire that motivates the appropriate actions. Another alternative is to postulate the existence

---

<sup>89</sup> Smith 1994, p. 74.

<sup>90</sup> Dreier 2000, p. 622. He does not explain why he finds the fetishism argument convincing.

of a plurality of moral desires with somewhat wide scope, such as a desire to promote justice, a desire to help those in need, a desire to help one's family and friends flourish, and so on.<sup>91</sup>

Shafer-Landau uses this approach to provide an alternative explanation for what is going on in Smith's political example. When Smith's friend convinces him to vote social democrat instead of libertarian, Smith's motivation "to support just political institutions, or more egalitarian political parties" leads him to become motivated to vote social democrat instead of libertarian.<sup>92</sup>

The close connection between moral judgments and moral motivation provides support for the intuition some have that there is a conceptual connection between the two. Caj Strandberg's recent work on morality and conversational implicature provides some interesting insights into this close connection from an externalist perspective. Strandberg has argued that this connection can be explained by drawing upon a Gricean theory of conversational implicature.<sup>93</sup> Strandberg argues that moral discourse has two purposes.<sup>94</sup> First, we use moral language to express our moral beliefs. If one states sincerely that  $\phi$  is morally wrong then one is expressing their belief that  $\phi$  is morally wrong. As a result, Strandberg's view is cognitivist. Second, moral language is used to influence peoples' behavior. If one states that  $\phi$  is morally wrong then this indicates that the person does not want that their audience  $\phi$ 's. Strandberg calls his view the dual aspect account and gives the following statement of it:

*The Dual Aspect Account (DAA):* A person S's utterance of a sentence of a type according to which " $\phi$ ing is wrong," conveys two things: (i) The sentence expresses, in virtue of its conventional meaning, the belief that  $\phi$ ing has a moral property. (ii) An utterance of this type of sentence carries a generalized conversational implicature, GCI, to the effect that S has a certain action-guiding attitude in relation to  $\phi$ ing.<sup>95</sup>

So when we make a moral statement like the one in the quote above, it is a part of the conventional meaning of that statement that we believe that  $\phi$  has a moral property (wrongness)

<sup>91</sup> See Shafer-Landau 1998 for an example of this kind of approach. Cuneo 1999 pursues a similar line of thought.

<sup>92</sup> Shafer-Landau 1998, p.356.

<sup>93</sup> See Strandberg 2011, 2012, and 2013. I draw mainly on the first two papers.

<sup>94</sup> Strandberg 2011, sec. 4 and 2012, sec. 8.

<sup>95</sup> Strandberg 2012, p. 101.

which corresponds to the predicate ('wrong') that we apply to it. In addition to its conventional meaning, the utterance carries a generalized conversational implicature (GCI) that indicates the speaker does not want that her audience  $\phi$ 's.

According to Strandberg, the conversational implicature is generalized rather than particularized because in most situations in which one utters ' $\phi$ ing is wrong' it will carry the implicature. That is, there are not special features of context that are necessary in order for the statement to have the implicature in question. The implicature is conversational rather than conventional because it is not retained in some contexts, such as embedded within certain conditionals. Strandberg states, "a person's utterance of the sentence 'If it's wrong to hit one's children, it's wrong to hit other people's children' does not necessarily convey any attitude."<sup>96</sup> Strandberg claims that it is also possible to calculate and cancel the implicature.<sup>97</sup>

It is not my intention to evaluate Strandberg's claim that the action-guiding aspect of moral *statements* like, 'It is right to  $\phi$ ', is due to a GCI rather than being part of the conventional meaning of the statement though I am sympathetic to his project. However, there is an important distinction between moral statements and moral judgments. The former are primarily public and shared whereas the latter are primarily internal (to an individual) and private. Strandberg is sensitive to this difference and claims that the dual aspect account can help us to make sense of the intuitions that underlie internalism.<sup>98</sup> The close connection between the conventional meaning and the GCI leads us to expect the two to co-occur. If I state sincerely that it is wrong to  $\phi$  and it is part of my purpose to get you to refrain from  $\phi$ ing then it is plausible that I myself am motivated to not  $\phi$ . We also tend to confuse the conventional meaning of a statement with its GCI. If this is so, it seems plausible that we might mistakenly think that the action-guiding nature of moral statements is a part of the conventional meaning.

---

<sup>96</sup> Strandberg 2012, p. 95.

<sup>97</sup> Strandberg 2012, sec. 12.

<sup>98</sup> Strandberg 2013, pp. 47-49.

Yet, it is not clear that we should treat judgments and statements the same here. Strandberg claims that internalists tend to deal with utterances rather than judgments.<sup>99</sup> I do not think he is correct but this is not my primary concern. It seems to me that our evidence for internalism is initially first-personal and then confirmed, to some extent, by our conversations with others. Consider. When I make a moral judgment of some sort it does seem to me that I have some motivation to act in accordance with that judgment. Even when there are other motivations guiding my behavior it seems possible that the motivation to do what I judge is right is still there but is being overridden by stronger, competing motivations. Though I think it is far from obviously true that people are morally motivated in all such cases, the claim does have some initial plausibility. Suppose however that when speaking to others about moral judgments I discover that most people typically feel no motivation to act in accordance with judgments that they make. In such a case, I would be inclined to judge that my reactions were atypical.

Another issue is that it seems that the action-guiding nature of moral judgments and moral statements can come apart in a way that is revealing. Suppose that I make a judgment that it is wrong for people in affluent countries to eat meat and that this judgment is accompanied by some motivation to adopt a vegetarian lifestyle. On the other hand, suppose that I use an utterance of 'It is wrong for people in affluent countries to eat meat', in a presentation that I am giving. In this case, the utterance may very well not have any action-guiding purpose and it may not even be a claim that I agree with—I might be using it for purely illustrative purposes. It seems at least possible that moral judgments could always be accompanied by motivation even if the action-guiding aspect of moral statements constitutes no part of their conventional meaning. This may be simply a fact of our psychology. Arguably, the reverse situation may hold as well. I might also make the above statement with the intention of influencing peoples' behavior even though I have no inclination to act in accord with it myself. For example, suppose we are having a pizza party and there is some pizza with pepperoni and

---

<sup>99</sup> Strandberg 2013, p. 47.

sausage, which I prefer, and some pizza that is vegetarian. I make the statement with the hopes of getting people to choose the vegetarian option, leaving more pepperoni and sausage slices for me.

## 2.7 Summary and preview

In this chapter I have examined the critique of moral externalism Smith gives in chapter three of *The Moral Problem*. His overriding point is that moral motivation is *de re* instead of *de dicto* because *de dicto* motivation would be fetishistic. Further, he maintains that the externalist needs to embrace *de dicto* motivation in order to explain the reliable connection between making a moral judgment and being motivated accordingly. I argued that much of the work that Smith does in this chapter is flawed—the color analogy does not make the point that he wants it to make, the political example does not show what he wants it to show, and *de dicto* motivation need not be fetishistic.

At this point I have only shown that the following conditional is true; *if de dicto moral motivation exists then it is not fetishistic*. In chapter four I will present a case which I believe will show that the antecedent of the conditional is indeed true. Before turning to this project, I will examine Smith's solution to the moral problem. Smith's solution relies on a crucial distinction between motivating reasons and normative reasons. I will argue that Smith's solution fails because there is good reason to think that there are no such things as *moral* normative reasons.

## Chapter Three

### 3.1 Introduction

In this chapter I will begin by critiquing Smith's solution to the moral problem. I will argue that while Smith provides a strong structural solution to the moral problem he fails to provide a substantive solution. Next I examine his critique of the naturalistic solution to the moral problem that might be provided by appealing to what are sometimes called desires. I will show that his case against them is not conclusive. I consider what is required to establish that a particular kind of mental state exists in anticipation of the work to come in chapter four and chapter five.

### 3.2 Smith's solution to the moral problem

By the end of chapter five of *The Moral Problem*, Smith has completed his case for the existence of normative reasons. Further, it appears that Smith's theory can provide a solution to the moral problem. First, our moral judgments can express beliefs about objective matters of fact. What we have normative reason to do is a reflection of what we would desire to do if we were fully rational. Second, we can explain the fact that if we have a moral belief then, *ceteris paribus*, we will be motivated to act in accordance with such a belief. This is because the belief about what we would rationally desire creates desires. Finally, we can respect the Humean intuition that desires and beliefs are distinct existences and that it is a desire that motivates moral action (coupled with an appropriate means-end belief).

Smith closes the fifth chapter with the following telling statement: "I also consider (in the next chapter) the substantive question whether there are any normative reasons. To anticipate: I argue that we have good reason to believe that there are."<sup>100</sup> Smith has given us a structural solution to the moral problem but, since moral reasons constitute a subset of our normative reasons then, if there are no moral normative reasons, this will pose a serious problem for

---

<sup>100</sup> Smith 1994, section 5.11, p. 181. All future references to Smith in this section will be from Smith 1994, unless otherwise noted.

Smith's solution. I will argue in this section that Smith's view does not give us good reason to think that there are moral reasons.

Smith suggests that in order to identify the proper subset of normative reasons which are also moral we should look to platitudes about morality. He suggests the following two: "Right acts are often concerned to promote or sustain or contribute in some way to human flourishing' (and) 'Right acts are in some way expressive of equal concern and respect.'"<sup>101</sup> Each of these platitudes is broadly drawn; the inclusion of the phrase 'in some way' guarantees that. Still, I wonder if it might be possible to raise a relativistic objection that not all cultures would accept both of these platitudes. For example, some cultures seem to be less concerned with 'equal concern and respect' than others. However, I do not intend to press this objection here. If Smith can give us a reasonable explanation of how a fully rational person settles on the most central of our moral norms this would go a long way toward answering such a relativism objection. Perhaps societies which do not show equal concern and respect for different persons are acting irrationally. Unfortunately, I do not think Smith's account can do this.

In section 6.4, Smith considers an objection to rationalism raised by both Foot and Gilbert Harman. The main point of the objection is that a person who behaves immorally can be accused of immorality but not always irrationality. Oftentimes it is irrational for a person to behave immorally because there are various negative consequences from immoral behavior. However, we have all found ourselves in situations in which we (apparently) could profit from an immoral behavior in which there was little or no chance of being discovered. In a passage quoted by Smith, Harman argues that a criminal can be completely rational even though he lacks any kind of concern for the harm that he may cause to others.

Smith claims that it is easy for him to meet the challenge posed by Harman and Foot. I admit to being confused by Smith's response so I will try to spell out what I think he must mean. He says that "the successful criminal must begin his deliberations from some evaluative premise

---

<sup>101</sup> Smith, section 6.1, p. 184.



or other.”<sup>102</sup> So, Smith seems to be suggesting that we start from the criminal’s attempt to justify his behavior. The start of his career may very well have not included any attempt to justify his behavior.<sup>103</sup> He committed crimes because he wanted the benefits that could accrue and believed that he could acquire such benefits. But now, in his justifications, he believes that “he has a normative reason to gain wealth no matter what the cost to others.”<sup>104</sup> Smith disagrees with the idea that the successful criminal, or any of us, possesses the normative reason above.

I quote at length:

For as we have seen, this (having a normative reason) is equivalent to the claim that fully rational creatures would want that, if they find themselves in the circumstances of the successful criminal, then they gain wealth no matter what the cost to others. And the successful criminal’s opinion notwithstanding, it seems quite evident that we have no reason to believe that this is true. Fully rational creatures would want no such thing.<sup>105</sup>

One thing that begs for an explanation in the above passage is why it is “quite evident” that the successful criminal lacks the appropriate normative reason. Smith states in the next paragraph that it is not the fact that just about everybody disagrees with the successful criminal that shows that he is wrong. But in the paragraph after that one he says that so many people disagreeing with him (the successful criminal) ought to give the successful criminal pause. Why is he right and so many other people wrong? Smith suggests that such a person “suffer(s) from the all too common vice of *intellectual arrogance*.”<sup>106</sup>

Certainly Smith does not want to appeal to just common opinion. This would be a blatant example of the ad populum fallacy. It is true that we appeal to intuitions when doing moral philosophy and it is also true that these intuitions tell us that it is not morally acceptable to gain wealth regardless of the cost to others but, for many of us at least, they do not tell us that it

---

<sup>102</sup> Smith, section 6.4, p. 194.

<sup>103</sup> The type of reasoning that the person would go through would seem to be the same regardless of whether he or she engages in it when trying to determine whether to embark on a career of crime or after having become a criminal though it seems some of the non-moral considerations would differ.

<sup>104</sup> Smith, section 6.4, p. 194.

<sup>105</sup> Smith, section 6.4, p. 195.

<sup>106</sup> Smith, section 6.4, p. 196. The italics are Smith’s.

is irrational to do so. What Smith needs to provide is an example to show why rational people would be led *by their rationality* to reject the successful criminal's guiding principle.

It may be helpful in trying to apply Smith's solution to start from a non-moral example. I am thinking about whether to work on my dissertation this evening or to spend the evening watching television. I am tired and desire to watch television. I ask myself what my fully rational self would desire in these circumstances. (Of course, not being fully rational I will only be able to do my best to figure out what my fully rational self would desire.) I decide that my fully rational self would not desire to spend the evening watching television since this is self-destructive and therefore not in my self-interest. If I work on my dissertation I will be that much closer to finishing it. When I earn my Ph. D I will have a better chance of securing a tenure track position and will get a large increase in pay. Though I still may be tired, I also remind myself that I enjoy doing philosophical work. Once I start engaging with the material I will probably enjoy myself and be happy that I did not spend the evening watching television. Given these considerations I believe that, were I fully rational, I would desire to work on my dissertation more than to watch television. If Smith is correct, and if my reasoning is correct, then anyone else in the same circumstances will also have normative reason to work on their dissertation.

It would be helpful if Smith actually took us through the reasoning that he believes the fully rational person would go through to see if she had normative reason to *gain wealth no matter what the cost to others*. It is not sufficient to appeal to our intuition that most of us think that we do not have moral reason to do so. Smith's theory should show that we do not have such a reason. As far as I can tell there are two basic ways that the successful criminal's reasoning might go.

1. I am a successful criminal and have been gaining wealth no matter what the cost to others. However, I wonder if I have normative reason to do so. Though I have been successful up to this point, it seems plausible that I will make a mistake at some point. It is hubris to think otherwise. Even if I do not make a mistake and am meticulous in my planning of my crimes there are things which occur for

which I cannot plan. I may be unlucky and get caught. If this happens, I will probably go to prison, perhaps for a long time. And I definitely do not desire to go to prison. I do have skills that I can employ doing honest work. Further, if I “go straight” then I will not be constantly looking over my shoulder. The stress of being a criminal interferes with my ability to sleep and may eventually take a toll on my health. Given these considerations I believe that, if I were fully rational, I would desire to abandon my criminal career and pursue honest work.

According to Smith, once the successful criminal has come to such a realization his belief about what he would desire if he were fully rational can create a desire to change his vocation.

Whether this desire is ultimately effective depends upon other desires the person has. Now consider the second way the criminal’s reasoning may go.

2. I am a successful criminal and have been gaining wealth no matter what the cost to others. However, I wonder if I have normative reason to do so. Though I have been successful up to this point, I also should think about the people that I have hurt. Some people have been harmed physically through my use of violence whereas others have presumably been emotionally traumatized when they came home to find their homes ransacked. Some people have been financially harmed because I stole property from them. Others may have had insurance but I know that my stealing raises the rates of everyone, at least a little. Further, every time someone reports one of my crimes a police officer must investigate. This must be passed on to the taxpayers in higher tax bills. Overall I believe that, if I were fully rational, I would desire to abandon my criminal career and pursue honest work.

As with (1), the reasoning employed in (2) can explain why the criminal would form a desire which would motivate him to change vocations. The main way that (2) differs from (1) is that (2) employs moral considerations whereas (1) employs prudential considerations. On its face, neither form of reasoning is implausible. We do frequently offer prudential reasons or moral reasons for our actions. However, if the criminal reaches the conclusion that he does not have normative reason to gain wealth no matter what the cost to others by employing the reasoning in (1); this cannot provide a sufficient account of the kind of moral reasoning that Smith is after. Intuition tells us that he is not engaging in actual moral reasoning but rather prudential reasoning. Each consideration offered shows how the criminal would benefit from abandoning his life of crime.

Smith might counter that anyone who finds himself or herself in the criminal's circumstances would have normative reason to abandon his or her criminal pursuits. This seems unlikely to be true. The problem is that if we alter the criminal's circumstances he now has normative reason to continue his life of crime. Suppose the criminal does not fear going to prison and/or is not stressed by his life of crime—at least not any more than he would be doing honest work. Now he does seem to have normative reason to pursue a life of crime because his fully rational self would advise him to continue his criminal ways. If (1) is what Smith has in mind it will not answer the challenge presented by Foot and Harman. Our intuitions tell us that fearlessness in a criminal does not change his actions from being wrong to right.

Suppose (2) is what Smith has in mind. (2) takes the various harms done to the victims of his crimes as reasons for the criminal to reconsider his vocation. But this prompts the question, "Why should the fact that action  $\Psi$  causes  $\mathcal{E}$ , where  $\mathcal{E}$  is taken to be something that is morally relevant, be taken as a reason for or against  $\Psi$ ?" Referring to (2) above, suppose  $\mathcal{E}$  is emotional trauma. It seems that the criminal should ask himself, "Do I have normative reason to cause emotional trauma to others when doing so can benefit me in some way?" The criminal will have to give either prudential or moral reasons. If he gives the former then, as discussed above, even if he finds that he does not have normative reason to cause others emotional trauma this will be at best a hypothetical reason. If we change his circumstances then he will come to the opposite conclusion. On the other hand, if he gives moral reasons and comes to the conclusion that he does not have normative reason to cause emotional trauma to others when doing so can benefit him, we must ask if his moral reasons are justified. Thus, we appear to be on the cusp of a regress. It does not appear that we can ever find our way out of this pattern of offering reasons and then seeking justification for those reasons. Each new plausible justification will offer moral reasons which will themselves stand in need of justification.<sup>107</sup>

---

<sup>107</sup> Some might think that the criminal might cite the guilt that he feels when he thinks about what he has done and that this might provide a solution to the problem. We must ask why this guilt is a motivating

How might Smith try to respond to this objection? He could try to argue that moral reasons receive their rational justification from other moral reasons. Trying to spell out this option in detail would take considerable time and would lead me away from my main project but I will raise what I believe to be one important drawback to such an option. It is difficult to see how we would be able to eliminate objectionable moral reasons. Suppose that one wanted to argue that it was immoral for women to have sexual intercourse outside of marriage. In a secular western culture one might find that there are not many reasons that others would be likely to accept. However, in some communities in which prohibitions against sex outside of marriage for women are common, one would be more likely to find support. Perhaps through appealing to notions of feminine virtue or beliefs about what is in a woman's best interest or what it is a normal for a woman to desire we could find the support we needed to argue for the conclusion that women having sex outside of marriage is immoral. One might counter that the argument is not cogent because one or more of the reasons offered in favor of the conclusion are false. Though I am sympathetic to this strategy, it does not appear to be Smith's.

A second strategy is to argue that the reasoning the successful criminal would go through if he were reasoning correctly would cite both moral and prudential reasons. However, it does not appear that this will be a viable option. Imagine that the successful criminal gives such an argument and comes to the conclusion that he does not have normative reason to gain wealth regardless of the cost to others. Either the conclusion will turn on prudential considerations or it will not. Suppose that it does. In this case, the conclusion of the argument will be relative to the circumstances of the criminal and the conclusion will not follow if we change the criminal's circumstances in the requisite ways. That is, some criminals will wind up

---

factor for the criminal. If the criminal cites the unpleasant experience of feeling guilt as a reason to change his ways then his reasoning is prudential and falls prey to the concerns raised above. If he lacked the guilt then he may lack the normative reason to change his ways and this cannot be what makes the difference. On the other hand, if his guilt is taken as a sign that he has behaved immorally we must ask if his guilt is appropriate as people often feel guilty when they should not. At best, his guilt is a sign that his behavior needs further investigation to determine if it is justified.

having normative reason to gain wealth regardless of the cost to others. Let us now consider the case in which changing the prudential considerations will not change the conclusion of the successful criminal's reasoning. In this case, it is the moral considerations that are doing the real work and the prudential considerations can be dismissed. In this case we must ask if the moral considerations themselves are justified, as we did with (2) above. A regress threatens once more.

A third option is to try to argue that the truth of at least some moral propositions is self-evident. On this view, we can realize that such propositions are true simply from reflecting on the proposition itself rather than inferring its truth from other propositions.<sup>108</sup> A proposition's being self-evidently true is not the same thing as being obviously true. A good deal of reflection may be needed to recognize that it is true. Some mathematical truths may be like this. In this context, reflection is generally understood to play a role analogous to that which perception plays in some of our beliefs about the external world. Looking at my desk I can just see that a coffee cup is sitting on the desk. A believer in self-evident moral propositions might argue that our knowledge that (e.g.) making money regardless of the cost to others is just morally wrong comes about in a similar way.

Regardless of the merits of such a strategy, it is clearly not an option for Smith. First, Smith clearly rejects this kind of non-naturalist strategy. He claims that the non-naturalist is committed to the claim that moral properties and natural properties are coinstantiated and that the defenders of such a view "owe us an account of how we come to have knowledge of" the relations between moral and natural properties.<sup>109</sup> He concludes that they have failed to do this. Second, Smith's account does not seem to be compatible with this kind of strategy. According to Smith, we do more than reflect on a proposition like "People have reason to gain wealth regardless of the cost to others" to determine its truth. We must ask ourselves if we would

---

<sup>108</sup> It is possible that some moral beliefs may be both self-evident and knowable through inference. See e.g. Shafer-Landau, p. 248 on this point.

<sup>109</sup> Smith, section 2.4, p. 25.

desire to gain wealth regardless of the cost to others if we were fully rational. This will presumably require us to speculate about what we would be like were we fully rational and the truth regarding whether or not we would have the desire in question will be inferred from the propositions that we think would be true of us if we were fully rational.

Finally, Smith could point to the level of moral agreement that exists between rational people as evidence that rational people will agree on these issues. He raises three points which he believes are relevant. First, there is actually much more moral agreement than is commonly recognized and these wide areas of agreement can be seen in thick moral concepts. Second, we recognize that we have made significant moral progress in the areas of “slavery, worker’s rights, women’s rights, democracy, and the like.”<sup>110</sup> Third, where there is still entrenched moral disagreement it is often because something like a religious authority is inhibiting rational discussion with others which could lead to our resolving these issues.

While I share Smith’s intuitions concerning moral progress and what frequently impedes it, I do not think that he has made his case that the moral progress we have seen stems from a priori rational deliberation. David Sobel raises the following objection to Smith’s claim:

Smith’s historical case about a tendency towards convergence will have to be genuinely historical. It will have to persuade us of the crucial role of facts, logic, and reason in explaining the history of convergence and the secondary role of force, guile, and a shared thick moral vocabulary.<sup>111</sup>

Smith needs a string of cases in which (1) convergence occurred, (2) it occurred for the right reasons, and (3) it occurred between radically different cultures.<sup>112</sup>

Sobel is making the point that in order for Smith’s case to be convincing Smith needs to provide evidence that the progress came about as a result of rational deliberation. Sobel makes the following claim:

Much of the historical moral argumentation that has actually managed to produce consensus has been (1) factually and logically imperfect, (2) addressed to those poorly positioned to object, (3) addressed to those who share substantial

---

<sup>110</sup> Smith, section 6.3, p. 188.

<sup>111</sup> Sobel, p. 146.

<sup>112</sup> Ibid.

common moral vocabulary, moral education, and cultural identification, and (4) offered by those who are persuasive for reasons other than the cogency of their position.<sup>113</sup>

It is not clear that argumentation has been the driving force behind our coming to agreement on the kinds of issues that Smith has in mind. Further, the fact that almost all rational people agree on something would not suffice to show that the agreement must be in virtue of their being rational. There may be irrational factors that explain why there is near unanimous agreement. For example, it may be because almost all rational people also happen to have been taught moral rules in their childhood's that prohibit the acquisition of wealth regardless of the cost to others. Having become deeply engrained these rules elicit unanimous agreement among rational people but it may be that the reason that these rules elicit agreement is that the violation of these rules elicits a negative emotional reaction. Both *de dicto* and *de re* explanations are available to explain this agreement. It may be that these rational people have a general concern about doing the right thing. (As we saw in the last chapter, *de dicto* explanations are not implausible.) It is also possible that these rational people cannot think of (e.g.) gaining wealth regardless of the cost to others without also being motivated to be opposed to it. This however does not show that these rules are agreed upon *qua* rational rules. At different times and different places, otherwise rational people appear to have endorsed slavery, been opposed to interracial marriage and so on. The violation of (e.g.) a rule against interracial marriage may very have elicited the same agreement.

In conclusion, even if Smith has made a convincing argument for the existence of normative reasons in general he has not given us good reason to think that there are any normative moral reasons. Though Smith's account does seem to provide a structural solution to the moral problem this solution comes at a high cost, the threat of moral nihilism. Given these considerations I believe that we must reject Smith's solution to the moral problem.

---

<sup>113</sup> Ibid. Ironically, Sobel fails to provide any historical evidence for the claims in the above quote. However, this does not weaken his main point that Smith does not provide historical evidence to show that moral agreement has come about through rational deliberation.



### Section 3.3 The method of reflective equilibrium

I suspect that Smith would respond to the criticisms in the previous section by saying that I have missed the point in a rather serious way. The concern can be put in a couple of different ways. First, it may be argued that I have been assuming that practical rationality must be instrumental but that reasoning in a practical way about morality is not concerned with instrumental reasoning; it is reasoning about final ends. Put another way, when we reason about morality we are reasoning about categorical rather than hypothetical imperatives. Suppose for illustrative purposes that Kant is right and that we can derive from the categorical imperative a duty to always refrain from making false promises.<sup>114</sup> Since I know that I have duty to refrain from making false promises and that I am aware of this duty through the use of reason, I would be behaving irrationally if I make a false promise even if I desire to do so. Reason tells me that I ought not to make a false promise regardless of the desires I happen to have. Now Smith does not appeal to the categorical imperative to justify our moral platitudes but rather mentions the method of reflective equilibrium. Smith argues that we employ the method of reflective equilibrium to see if our desires are systematically justifiable in a way that is similar to the way in which a set of beliefs could be systematically justifiable.<sup>115</sup> I will argue in the following that Smith's argument fails but to see this it will be useful to take a closer look at the method of reflective equilibrium, hereafter MRE.<sup>116</sup>

The MRE receives its name from John Rawls and is featured prominently in his work.<sup>117</sup> To apply the MRE, we can start by examining our specific and general beliefs about ethics. It is

---

<sup>114</sup> And that the categorical imperative is true as well.

<sup>115</sup> See Smith 1994, section 5.9, especially pp. 155-161.

<sup>116</sup> In the paragraphs to follow, I sketch a picture of the nature of reflective equilibrium. However, it seems that not everyone would agree with all the details because not everyone agrees on the details of MRE. However, these disagreements should not be important here as long as an accurate general picture emerges.

<sup>117</sup> See e.g. Rawls 1951, 1971, and 1974-1975. I will draw primarily upon Rawls 1974-1975 and Daniels 1979 in my explication of the MRE.

natural to think of the specific beliefs as our judgments about particular cases and our general beliefs as the moral principles that we hold, though some principles may be more general and others more specific. We can also have general beliefs about morality that do not take the form of principles, such as that the moral rules are the same for everyone. To apply the method, we move back and forth between our specific and our general moral judgments, making adjustments when necessary to achieve as much consistency as possible.<sup>118</sup> A simplified example will be helpful. Suppose I am gathering together my moral beliefs and come up with the following three:

1. It is wrong for me to hit my brother.
2. It is wrong for me to pull my sister's hair.
3. Moral rules are the same for everyone.

Seeing as I believe (3) and I cannot think of exceptions (e.g. I do not believe that it is okay for you to pull your sister's hair), I decide that (1) and (2) can be expressed in more general terms as (4) and (5).

4. It is wrong for anyone to hit their brother.
5. It is wrong for anyone to pull their sister's hair.

This adjustment makes sense as my belief in (3) is robust whereas my belief in the specificity of (1) and (2) was weak. As the specificity of (1) and (2) may have been taken to imply, colloquially speaking, that these rules did not apply to everyone it made sense to replace them with (4) and (5). Now, looking at (4) and (5) may lead me to formulate the general principle below.

6. It is wrong to hurt a sibling.

I now have (3)-(6) in my collection of beliefs ((4) and (5) encompassing (1) and (2)) and it may occur to me that I believe that it is not okay to hurt non-siblings as well, which may lead to a further revision in my beliefs and so on.

---

<sup>118</sup> Sometimes the word 'coherence' is used to describe what we are after in MRE but, as far as I can tell, writers usually have logical consistency in mind.

Though this procedure has a certain amount of intuitive appeal regarding how one might modify their moral beliefs, it is not without problems. To address these, Rawls makes an important distinction which becomes explicit in Rawls (1974). Rawls distinguishes between narrow and wide reflective equilibrium. A problem with MRE as sketched so far is that judgments and principles may be unduly influenced by contingent factors such as one's social class, culture, and so on.<sup>119</sup> Imagine two persons, A, who has been raised by devoutly utilitarian parents and B, who has been raised by devoutly Kantian parents. A has been taught that the principle of utility is the most general principle and must take priority over all more specific principles as well as judgments about specific cases and B has been taught that the categorical imperative gives rise to more specific duties which then dictate our behaviors in specific situations. A and B must then make a judgment about a specific situation in which making a false promise will result in greater expected utility than any other available course of action. A's "intuition" about the right thing to do in the case can be expected to differ from B's, yet both are justified within their respective belief sets. Further, there appears to be no obvious reason why there would be a difference in the consistency, simplicity or generality in A or B's belief sets that would tip the scales in favor of one rather than the other.<sup>120</sup>

One way to try to resolve this problem is to strive for wide reflective equilibrium rather than narrow reflective equilibrium.<sup>121</sup> Daniels says that we can do this by including background theories in our reasoning that are "more than reformulations of the same set of considered moral judgments involved when the principles are matched to moral judgments."<sup>122</sup> He goes on to argue that Rawls succeeds in showing that justice as fairness is more plausible than utilitarianism.<sup>123</sup> If a theory that is generated by the MRE is more plausible than a second theory

---

<sup>119</sup> Daniels 1979, p.257.

<sup>120</sup> Generality and simplicity are also often taken to be points that count in favor of a theory that emerges in reflective equilibrium.

<sup>121</sup> Hereafter MRE will stand for the method of wide reflective equilibrium unless otherwise noted.

<sup>122</sup> Daniels 1979, p. 259.

<sup>123</sup> Daniels 1979, p.260-261 and Daniels SEP.

with which it might be compared, this speaks in favor of the first theory. Another way of interpreting the phrase 'background theories' is as including theories that are not primarily about morality, such as various scientific theories. Consistency with such theories speaks in favor of a moral theory and inconsistency speaks against it.

The MRE is not without its critics. One concern that has been raised is whether the correct application of the MRE can show moral truth. That is, suppose that justice as fairness is chosen in MRE over all other theories. Does this show that justice as fairness is true? There seems to be good reason to doubt this. Rawls himself often does not seem to have thought of the theory in this way. Rawls says that we can use the MRE to "understand our sense of justice" by organizing our beliefs about justice in a systematic way so that we know what our principles are.<sup>124</sup> Rawls does sometimes indicate that he thinks that the MRE could show that a theory is true. At one point he says the following: "I first discussed the method of reflective equilibrium and suggested that the question as to the existence of objective moral truths seems to depend on the kind and extent of the agreement that would be obtained among rational persons who have achieved, or sufficiently approached, wide reflective equilibrium."<sup>125</sup>

Two main problems arise for Rawls' claim. First, it seems unlikely that such a convergence would emerge. If initial starting points differed significantly for different individuals, and given different cultural backgrounds we should suspect that they would, then it is seems probable that their end points would differ significantly as well.<sup>126</sup> Consider again the example given above of two people, one raised as a Kantian and the other as a utilitarian. It is hard to see how they are going to converge on questions about exceptions to moral duties in cases where expected utility is maximized by violating those duties. Starting intuitions can be shaped by our upbringing and it is not clear that any non-moral theory will sway us one way or the other. Second, even if convergence does occur this would not be sufficient to show that the theory is

---

<sup>124</sup> Rawls 1971, p.41 but all of section 9 is relevant.

<sup>125</sup> Rawls 1974-1975, p. 21.

<sup>126</sup> See Kelly and McGrath (2010), p. 337-339.

true. K. Kappel raises a serious concern for the prospect of using the MRE to produce moral truth even in cases of convergence.<sup>127</sup> Suppose that the application of MRE leads us to converge on a moral theory, T, and further that T happens to be true. The defender of MRE needs to show that the MRE led us to converge on T *because* T is true and not for other reasons. Kappel states that “if we claim that the convergence obtained through the use of MRE is best explained by truth, then it seems we need to assume in advance that MRE works essentially be tracking truth.”<sup>128</sup>

I have discussed serious concerns others have raised about whether the MRE can be used to show that a particular moral theory is true. Setting this concern aside, would Smith’s moral rationalism be shown true by the MRE even if MRE could show truth? I do not believe it would. In chapter two of *The Moral Problem*, Smith considers expressivist moral theories. The expressivist solves the moral problem by denying thesis one and maintaining that the seemingly factual nature of moral claims is illusory. Moral claims do not express beliefs but rather some sort of pro-attitude toward the behavior in question. As a result, the connection between a moral judgment and motivation is explained.

Smith examines A. J. Ayer’s claim in *Language, Truth and Logic* that moral claims can be neither descriptions of natural states of affairs nor non-natural states of affairs. Smith claims that Ayer’s main objection to non-naturalism succeeds and that non-naturalism is probably false but that his objection to naturalism fails because the naturalist need not be giving a description. Ayer has not shown that moral truths cannot be reconciled with a broader naturalism in which we do not try to give reductive moral truths.<sup>129</sup> Assuming that everything that has been done in chapter two of *The Moral Problem* is successful, Smith has shown that the expressivist project

---

<sup>127</sup> Kappel 2006.

<sup>128</sup> Kappel 2006, p. 138. In chapter five of this dissertation I argue that making a certain kind of ontological assumption about morality can help virtue ethics to make progress in meeting this requirement.

<sup>129</sup> Smith, section 2.14.

has not succeeded and that there *might be* moral truths which are natural in the wide sense of the term, meaning that they do not conflict with a broadly naturalistic view of the world.

In finishing up a critical discussion of Ayer's version of expressivism, Smith says that "we therefore leave Ayer's argument with a promise, a promise that we will indeed be able to provide an analysis of our moral concepts in summary-style, non-reductive terms, an analysis that does indeed vindicate the descriptive form of moral judgments."<sup>130</sup> What Smith is promising is an account that shows that moral judgments express statements, some of which are objectively true. That is, he is looking to vindicate the intuition that supports our belief in thesis one of the moral problem. Having ruled out the possibility that valuing is a type of desiring, Smith's preferred solution says that valuing is a type of believing. To value X is to believe that we would desire X if we were rational and fully-informed. Smith claims that "by far the most important way in which we create new and destroy old, underived desires when we deliberate is by trying to find out whether our desires are *systematically justifiable*."<sup>131</sup> He suggests that we do this by following something like the MRE and argues that a set of desires can be systematically justifiable in the same way that a set of beliefs can be systematically justifiable.<sup>132</sup>

There is some reason to think that Smith may be able to do this.<sup>133</sup> Consider how we might come to attribute the property of being fragile to a particular kind of object, such as a compact disc (CD) case. Having witnessed one or more CD cases break upon being dropped on a hard floor, I come to conclude that CD cases have the property of being fragile. Yet, I have no knowledge of what it is that makes them fragile. (Not all plastic things are fragile.) There may be more than one underlying property structure that can make something fragile. If so, we can say that the property of fragility supervenes on a disjunctive list of all the underlying

---

<sup>130</sup> Smith (1994), p. 59.

<sup>131</sup> Smith (1994), pp. 158-159.

<sup>132</sup> Smith (1994), pp.159-161. See also Smith's reply to Enoch in Smith (2007).

<sup>133</sup> In the next two paragraphs I draw heavily upon the work of Jackson (1998).

property structures that make something disposed to break easily when dropped. However, a neat reduction of fragility to any particular underlying molecular structure will not be possible.

Returning to the discussion of Smith in the previous section, we saw that Smith appealed to convergence as part of his justification for believing that there are moral truths. Suppose, I come to believe that there is such a thing as being fragile and that CD cases possess this property but everyone I talk to disagrees with me about this. I come to suspect that the CD cases that I have broken may have been of faulty construction. However, if my judgments converge with those of others who have experience with CD cases this will make me more confident that CD cases really do possess the property of being fragile. Further, we may feel confident in our belief regarding the existence of fragility of CD cases even if we all remain ignorant of the underlying bases of fragility.

Just as our convergence in the realm of the fragile is largely due to the underlying nature of fragile things we may come to think that our convergence in the realm of the moral is due to the underlying nature of the moral. If I make certain judgments—*slavery is morally wrong, helping those in need is the right thing to do, etc.*—and see that my judgments are the same as those that others make then I might be tempted to conclude that we are converging because we are latching on to some underlying truth in the same way that we are latching on to some underlying truth concerning the properties of CD cases.

Unfortunately, Sobel's criticisms seem to be on target here as well. We may not need to know why the convergence occurred but we should be relatively confident that the convergence did not occur due to reasons that are not concerned with the underlying basis to the moral realm. Though it is certainly possible to make judgments about fragility by being told by others that something is fragile, we also frequently make judgments about the fragility of CD cases by dropping them on the floor and seeing them break. It is reasonable to believe that fragility supervenes on the underlying physical properties of fragile things. In theory, we could give a disjunctive list of the different possible underlying property structures that make something

fragile. X is fragile if and only if X possesses  $p_1$  or  $p_2$  or  $p_3$  or . . .  $p_n$ . To rely on convergence we need to know why the convergence occurred. Sobel has given plausible alternative reasons for believing that the convergence is due to something else other than the underlying nature of morality.

Smith would have us believe that the convergence is due to the judgments of rational agents. Smith's solution works, I will argue, by making unwarranted assumptions about the nature of rationality. Smith says that "being fully rational is itself a summary notion" and that the idea of providing an analysis of rationality is "to capture, in summary style, a whole host of more specific platitudes about practical rationality."<sup>134</sup> So, it is a platitude of prudence that rational people try to advance their own interests, though perhaps not regardless of cost. For example, suppose I receive a paycheck from work and refuse to cash it, throwing it in the garbage instead. I do not have some compelling moral reason for refusing to cash the check, nor am I independently wealthy. In fact, I live from paycheck to paycheck. In this case, we would judge my behavior to be irrational. Smith argues that we also possess an even more specific group of platitudes concerning morality. He gives the following two: "'Right acts are often concerned to promote or sustain or contribute in some way to human flourishing', 'Right acts are in some way expressive of equal concern and respect' and the like."<sup>135</sup>

We can think of these moral platitudes and others like them as being subjected to the MRE but this does not show that they are the creation of practical rationality. To see this, consider a parallel argument that prudential truths are the creation of practical rationality. We can start by gathering our beliefs about prudential behavior. Consider the following four:

1. Prudent people do not engage in risky behaviors.
2. Prudent people do not do things to hurt themselves.
3. Prudent people pursue opportunities to advance their own ends.
4. Prudent people take advantage of opportunities that will benefit them greatly.

---

<sup>134</sup> Smith (1994), pp. 155-156.

<sup>135</sup> Smith (1994), p. 184.



Now we can subject these beliefs about prudence to the MRE. We may notice that (1) and (4) would lead to different judgments about particular cases and decide that (1) should be replaced by (5).

5. Prudent people do not engage in risky behaviors unless they may benefit greatly from doing so.

And so on. Again we would find that consistency, generality, and simplicity would be virtues of the account. Yet practical rationality does not create prudential truths but rather discovers them. Consider the following specific prudential truth: It would be imprudent to drink the gasoline in that glass (where the glass in question is made clear from the context). Surely a better explanation of what makes the claim true is that gasoline is poisonous if drunk and will cause serious harm to a human being rather than that a practically rational and fully-informed person would judge it to be true, even though such a person would presumably make this judgment. The practically rational and fully-informed person will be ideally situated to discover this truth. The truth of the statement itself is presumably reducible to facts about the chemical structure of gasoline, the biological make up of human bodies and the way the former acts on the latter. Prudential truths are thus reducible to other facts. The key point here is that reason can be used to discover prudential truths but does not create them.

Notice that we can use reason to discover prudential truths because we are applying our reason to a realm whose existence is independent of our thoughts about it. Further, these truths exist independently of an individual's desires. For example, I can judge that it would be imprudent of you to drink the gasoline in that glass without knowing anything about your desires. To put this in terms of reasons, let us say that you have reason to not drink the gasoline. Does this entail that you ought not to drink the gasoline? It depends on whether the 'ought' is read as an all-things-considered ought or only a prudential ought. If it is the latter then the answer is clearly 'yes' but if it is the former then the answer could potentially be 'no'. We would need to

determine if there were some conflicting ought claim and then try to weigh the two to see which takes precedence.

Smith may seem to avoid this problem by appealing to normative reasons. Whether a person ought to drink the gasoline depends on what that person's ideally-rational and fully-informed self would advise them to do. Yet no one seriously doubts that some prudential claims are objectively true though there are serious doubts about whether any moral claims are objectively true. The root of the problem seems to be that Smith is assuming precisely what he needs to show. He needs to show us that reason gives rise to the platitudes and that these platitudes, or most of them anyway, are objectively true or at least close to the truth *and then* we could subject them to the MRE. Of course, if we assume that the things that we normally assume are true about morality are true then the MRE will show that there are moral truths. This is because MRE is primarily a conservative process. Smith needs to show why we should think that what we input into the MRE is true to begin with.

To put this in terms of supervenience, it would be good if Smith could show that the moral platitudes supervene on the judgments that fully-informed rational persons would make. Yet the considerations in section 6.3 of *The Moral Problem* do not show this, for reasons that Sobel has made clear. Simon Blackburn raises a similar concern about supervenience applied to moral issues.<sup>136</sup> Blackburn claims that though it is possible that the moral supervenes on some set of natural properties, the moral theorist owes us an explanation of how this occurs. Blackburn favors a projectivist explanation. I will not evaluate his claims here. However, in chapter five I will make an attempt to show how one aspect of morality, the notion of rightness, can supervene on underlying natural properties. In the next two sections, I examine and evaluate Smith's claim that there can be no such thing as mental states with the properties of both beliefs and desires.

---

<sup>136</sup> See "Supervenience Revisited" in Blackburn 1993.

### 3.4 Smith's argument against besires

Given the above considerations we must conclude that Smith's attempt to solve the moral problem fails. Does this mean that no solution to the moral problem is available? Let us remind ourselves of the three theses which are independently plausible but are collectively difficult to consistently maintain.

1. Moral judgments of the form 'It is right that I  $\phi$ ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do.
2. If someone judges that it is right that she  $\phi$ s then, *ceteris paribus*, she is motivated to  $\phi$ .
3. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences.<sup>137</sup>

Smith resolves the tension by showing that (2) is not so strange. Having a belief that  $\phi$  is right can cause one to be motivated to  $\phi$  because believing that  $\phi$ -ing is right is what we believe that we would desire if we were fully rational. Beliefs about what we would desire if we were fully rational can then create desires. So, (2) is no longer mysterious and does not seem to clash with (3) because it is still a desire that is providing the motivating force.

We could skip the appeal to full rationality if moral beliefs were themselves the kinds of entities that had motivational force. One suggestion is that moral mental states are states that have the properties of both beliefs and desires. That is, they would both describe the way the world is and motivate us to make the world be a certain way. If such an answer were possible it would explain why moral beliefs motivate and we could reject the Humean theory of motivation, at least where it concerned moral motivation. J. E. J. Altham calls these states 'besires'.<sup>138</sup>

Smith considers and rejects this option in a nuanced and complex argument.<sup>139</sup> To start, Smith claims that such a state is impossible because "though it might sound like a coherent possibility . . . it isn't really." Continuing, "A state with both directions of fit would therefore have to be such that, both, in the presence of a perception it tends to go out of existence, and, in the

<sup>137</sup> Smith 1994, section 1.3, p. 12. All subsequent references to Smith made in this section will be to Smith 1994, unless otherwise noted.

<sup>138</sup> Altham 1986.

<sup>139</sup> Smith, section 4.7

presence of such a perception, it tends to endure, leading the subject who has it to bring it about that p.”<sup>140</sup> Smith does not elaborate on what he means in this dense and difficult passage. Here is how I understand him. Consider the following possible state the world could be in: Jones owning a Porsche sports car. If Jones desires to own a Porsche then he does not currently own one and is motivated to change the world so that he comes to own one. Suppose that Jones also believes that he owns a Porsche. If we imagine these two contents together in one mental state then the recognition of either part of this mental state would seem to eliminate the other part. The recognition of the belief that he owns a Porsche would presumably eliminate the desire for a Porsche whereas the recognition that he desires a Porsche would presumably eliminate the belief that he has a Porsche. The state seems to have parts that contradict themselves.<sup>141</sup>

As Smith recognizes, a different option is available for the besire theorist. The besire theorist can argue that a besire is a state that has different directions of fit with respect to different contents.<sup>142</sup> If we imagine that besires have a desire component and a belief component and that the desire component does not contradict the belief component then we can avoid the problem above. An example may be helpful here. Let us suppose that there is a helping besire. We can express the content of such a besire using the following bi-conditional:

I feel that I ought to help  $\equiv$  It is morally right to help people<sup>143</sup>

However, Smith thinks that there is a strong two-part argument against this type of solution. I will quote at length.

---

<sup>140</sup> Smith, section 4.7, p. 118.

<sup>141</sup> Someone might object that the two parts do not really contradict themselves. It might be the case that Jones already owns a Porsche and desires to buy another one. So he both believes he owns a Porsche and desires to own a Porsche. This problem can be resolved by re-describing the possible state in question as ‘Jones owning a second (or third or fourth) Porsche.’

<sup>142</sup> Smith, section 4.7, p. 118.

<sup>143</sup> The biconditional seems to be the most appropriate logical operator because the two contents must co-occur if they are two parts of the same state. The nature of the connection will be discussed below.

They (anti-Humeans) must claim that it is *impossible* for agents to be in a belief-like state to the effect that their  $\phi$ -ing is right not to be in a desire-like state to the effect that they  $\phi$ ; that the two cannot be pulled apart, not even modally.<sup>144</sup>

As I see it, the disagreement between Humeans and anti-Humeans as to whether we are just believers and desirers, or rather besirers as well, amounts to no more and no less than a disagreement about these modal claims.<sup>145</sup>

First, anti-Humeans must hold that it is not possible to separate the belief-like state from the desire-like state of a besire, even modally. Smith does not elaborate on what he means by the use of the term 'modally' in this context but it seems plausible that he means that the anti-Humean must hold that we cannot conceive of the desire-like part as existing separate from the belief-like part. Call this the conceptuality claim. It seems clear that we do frequently think and speak of moral beliefs and desires as existing separately. As a result, if the anti-Humean is forced to deny the conceptuality claim then she will be in serious trouble.

Smith goes on to argue that the anti-Humean is committed to the claim that a person who makes a moral judgment will be "motivated to act accordingly *simpliciter*"<sup>146</sup> rather than *ceteris paribus*. The reason to think this is because the belief-part exists in the same mental state as does the motivational part. If this is the anti-Humean position then Stocker's cases of persons suffering from depression, accidie, and other motivation-sapping conditions in which persons make moral judgments yet are not motivated to act accordingly will be something that the anti-Humean will struggle to explain. If what we call moral beliefs are actually besires then the Stocker-cases make it appear that it must be possible for the motivational side of the besire to vary independent of the content side of the besire. In addition, Nick Zangwill argues that if besires must be variable in this way then we have good reason to embrace a belief-desire explanation and reject the existence of besires.<sup>147</sup> A natural explanation of what happens in cases of depression is that our desires are dampened or eradicated completely whereas our

---

<sup>144</sup> Smith, section 4.7, pp. 119-120.

<sup>145</sup> Smith, section 4.7, p 120.

<sup>146</sup> Ibid. The italics are Smith's.

<sup>147</sup> Zangwill 2008a.

belief structures are left intact. (These disorders are, after all, called affective disorders.) Call Smith's claim that the anti-Humean is committed to the claim that moral beliefs will motivate simpliciter rather than *ceteris paribus*, the actuality claim.

Smith also believes that such a response works against John McDowell's anti-Humean view. John McDowell argues that virtuous persons conceive of moral situations in a certain way. In a passage quoted by Smith, McDowell argues that the virtuous person conceives of the fact that someone is shy and sensitive in such a way that it entails that the person be treated in a certain way. Though it is not unnatural to say of a virtuous person that she desires to, (e.g.) protect the shy and sensitive person from embarrassment, according to McDowell, this ascription is *simply consequential*. The reason in itself is sufficient to motivate. Smith thinks that McDowell's account fails for the same reasons as those considered above. If McDowell were correct and virtuous peoples' moral mental states "really were . . . besire(s), it would have to be impossible to break the belief-like part of that state apart from the desire-like part."<sup>148</sup> Yet we see that we clearly can, both conceptually in that we can think of the belief that it would be wrong to embarrass a shy and sensitive person as existing separately from the motivation to protect a shy and sensitive person, and that this separation actually occurs in the Stocker-type cases.

### 3.5 Evaluating Smith's argument against besires

Prior to evaluating Smith's argument against besires, it bears pointing out that the view of desires that Smith takes when arguing against the besire theorist seems a bit anemic. Consider a desire to eat a piece of cheesecake. Such a desire entails a number of beliefs such as that cheesecake exists, that it is the kind of thing that can be eaten, (probably) that it is tasty, and so on. Though it seems possible for a person to be in a state in which they are motivated with no idea what it is that they motivated to do, typically this is not how a desire works. Some

---

<sup>148</sup> Smith, section 4.7, p. 123.

sort of motivational state, or perhaps the disposition to have one in certain circumstances, *might* be a necessary condition for some state to be a desire but it could not by itself also be a sufficient condition. This is because the state would be functionally directionless. There must be some object toward which the desire is directed.

Dennis Stampe argues for an Aristotelian-style account of desires that holds that desires are representations of the apparent good.<sup>149</sup> On Stampe's view, desire is a form of perception in which we perceive bodily states. As a result, reasoning from desire starts from a perception, proceeds to an instrumental observation about how one could bring about what is perceived as good, and concludes with an intention. The interesting thing about Stampe's view for our purposes is that what he calls desires have the structure of a besire because they have both representational and motivational aspects. I will not critically examine the implications of Stampe's view here though I intend to return to it in future work.

Over the next few pages I will discuss the conceptuality claim and the actuality claim made in the previous section. I will argue that the conceptuality claim is false and that the actuality claim might be false. I will address the actuality claim first. The Stocker cases do present a problem for the besire theorist as both Smith and Zangwill point out. The fact that depression, accidie, physical tiredness and so on can leave our beliefs intact while eliminating the desires that frequently accompany such beliefs seems to speak strongly in favor of a belief-desire model. If a besire can at one point be equated with a mental state that has both a belief-like part and a desire-like part and, perhaps at a later time, could be equated with a mental state that has only the belief-like part then it would seem that the desire-like part is not essential to the besire. If this is the case then we appear to be dealing with a standard belief-desire model.

This last point shows why besires must be proportional rather than variable.<sup>150</sup> Following Zangwill, let us say that a besire is variable if it contains both a belief-like part and a desire-like

---

<sup>149</sup> See Stampe 1987 and 1990.

<sup>150</sup> The terminology is from Zangwill 2008a.

part and that the strength of the desire-like part can vary while the belief-like part remains the same and that the besire is proportional if the strength of the desire-like part cannot vary while the strength of the belief-like part remains constant.<sup>151</sup> If variable besires exist then they offer the promise of providing an explanation of the Stocker cases in which one seems to retain the belief that something is good but appears to have no desire to do what they believe to be good. In these cases, the (e.g.) depressed person presumably has some desire to what is do what he or she judges to be good but her or his motivation may be lessened to the point at which it is ineffective and perhaps even undetectable.

The variable besire option is unattractive. First, it seems that we are being asked to take it on faith that these people do feel some motivation even if they report that they do not. Zangwill gives examples in which otherwise normal people (i.e. those not suffering from conditions like tiredness, depression, etc., as is the case with the Stocker cases) seem to suffer from moral indifference.<sup>152</sup> These are not cases in which people seem to be suffering from a lack of rationality. Some will argue that the nature of morality is such that all rational people will be motivated by their moral judgments and that, despite how they may seem the mercenary and the early morning coffee deprived person are not rational. (These are two examples Zangwill gives.) Given their appearance of rationality, the onus is on the moral rationalist to offer us a convincing explanation of why, appearances aside, they are in fact irrational. Given the problems that I have raised for Smith's theory earlier in this chapter as well as in the previous chapter, the evidence seems to suggest that both the moral rationalist and the variable besire theorist are wrong.<sup>153</sup>

---

<sup>151</sup> However it may be the case with proportional besires that the two parts can vary together. For example, a weaker conviction that some act is (e.g.) the morally right thing to do in a situation may lead to a weaker motivation to do that act.

<sup>152</sup> Zangwill 2008b.

<sup>153</sup> In places Zangwill appears to take moral internalism and besire theory to be interchangeable but the latter should be taken as narrower than the former. A besire theorist regarding morality believes that there are mental states that contain both a desire-like part and belief-like part and that these parts are essential to the mental state. An internalist believes that it is in the nature of moral judgments to be accompanied by moral motivation though this motivation may be defeasible due to Stocker type cases



This seems to force the besire theorist into holding that the degree of desire will be proportional to the degree of belief. Suppose I have a besire which causes me to want to help when I believe that I am in a situation in which I can help. It may be that the more confident I am that this situation calls for help the stronger will be my motivation to help. What is not possible is for the desire-like part of the besire to fluctuate while the belief-like part remains constant. If besires would have to be proportional does this somehow show that besires cannot exist? Let us examine McDowell's view a little more closely.

Now I agree with Smith's analysis of the situation if we are required to classify *all* moral mental states as being either besires or belief-desire sets. Moral beliefs do seem to motivate only contingently, as Stocker points out. It does seem possible to have a moral belief without the accompanying desire. However, this fails to disprove the existence of besires. At most it shows that the belief that besires *exhaust* our moral mental states is mistaken. The reason why one person recognizes that someone is shy and sensitive and is thereby motivated to protect the person while a second sees that someone is shy and sensitive but feels no such motivation might be that the two people possess different kinds of mental states, as the following passages from McDowell suggest:

Someone who fails to act virtuously may, in a way, perceive what a virtuous person would, so that his failure to do the right thing is not inadvertent; but to insist that his failure occurs only because his appreciation of what he perceives is clouded, or unfocused, by the impact of a desire to do otherwise.<sup>154</sup>

One way to bring out the special nature of the conception (of virtue) is to note that, for Aristotle, continence is distinct from virtue, and just as problematic as incontinence.<sup>155</sup>

McDowell is arguing for an Aristotelian view in which there are two morally good states rather than just one. The first passage immediately above suggests that the virtuous person differs

---

and the types of indifference that Zangwill 2008b discusses. In brief, Smith believes that moral judgments are judgments about what we would desire if we were fully rational. The belief about what our fully rational self would desire then, *ceteris paribus*, causes the appropriate desire. Smith's theory should not be thought of as a type of besire theory.

<sup>154</sup> McDowell 1979, p. 334.

<sup>155</sup> *Ibid.*

from the (merely) continent person in that the virtuous person has a focus that the continent person lacks. The continent person may feel motivated to behave in a way that runs counter to what is the right thing to do. The continent person feels the pull both of what is right and of some other consideration—e.g. something that is in the agent's short-term interest. In contrast, the virtuous person is focused on what the right thing to do is and acts accordingly. There is no tug-of-war between short-term interest and the right thing to do not for the virtuous agent not because he fails to appreciate that something is in his short-term interest but rather because he does not allow such considerations to enter his mind.

However, I believe that a careful reading of McDowell suggests that the difference between virtue and continence as McDowell is describing them is not that they are different *mental* states but that the virtuous person has something that the continent person lacks, focus. On this reading, the mental state that the virtuous person shares with the continent person may both be *besires*, i.e. they may both have a descriptive and a motivating part. Lacking competing considerations, each person will do what is right. However, the virtuous person has the ability to maintain focus on the morally relevant factors in the situation whereas the (in)continent person does not have this ability and thus runs the risk of not doing the right thing whenever there are factors that could result in competing motivations.

However, what if persons actually possess two different kinds of moral states, one state that is intrinsically motivating, as internalists favor, and a second kind of state that does not include motivation and gets its motivation from an external source, as externalists believe? Much of the rest of this dissertation will be concerned with establishing the plausibility of such a position. One advantage of such a position is that it seems to provide a solution to the Stocker type cases. Suppose that a person, A, possesses both a *besire* whose content part is that it is good to give to charity during the holidays as well as a moral belief that it is right to give to charity. A's *besire* includes motivation to give to charity during the holiday season whereas A's moral belief is free of motivational force. However, suppose as externalists commonly do that A

also possesses a general desire to do that which is right, whatever that turns out to be. A usually gives money to charity during the holiday season but this year A is suffering from depression and feels no motivation to give to charity though she readily admits that she ought to give to charity.

The dual-state theorist has a ready explanation for what is occurring. A's depression has eradicated her free standing motivation to do what is right, whatever that turns out to be, *and* has eliminated her besire. A retains her free-standing moral belief about giving to charity but, lacking any source of motivation, she fails to do what she believes is right. It also helps to explain more common cases in which moral motivation is lacking but not absent completely, and it is clear that Stocker is concerned with these as well, as comes out in the following passage:

Lack of desire is commonplace. Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good.<sup>156</sup>

It is clear that Stocker is describing phenomena which affect the degree of moral motivation. It is not necessarily the case that a person who comes to feel (e.g.) generally apathetic about life goes from feeling a normal amount of moral motivation to feeling absolutely none.<sup>157</sup> A dual-state theory can explain this as well. Suppose A is experiencing a mild depression. In the case of a besire, the mild depression may both lessen the belief-part of the besire as well as the desire-part, making it less likely that the besire will result in action. Such a lessening would presumably be experienced by the agent as a lack of motivation but not a total absence. Such a depression, despair, etc., could also result in a lessening of A's free-standing desire to do what is right, whatever that may be. Again, the result would be less motivation and a lowered likelihood that A will do what is right. A dual-state theory can also help to preserve our common

---

<sup>156</sup> Stocker 1979, p. 744.

<sup>157</sup> What counts as normal here will be relative to the particular individual's character as well as other factors.

intuition that judgments about what is morally right are accompanied by moral motivation. If many of our moral mental states are *besires* then, absent depression, fatigue, etc., we will feel at least some motivation accompanying most of our moral judgments.

An objection to a dual-state theory immediately presents itself. First, the move appears to be *ad hoc*. Assuming that there are both moral beliefs and moral *besires* seems to offer a solution to the problem presented by the Stocker-cases but what independent reason do we have to accept such a theory? Lacking independent reasons to adopt a dual-state theory, Occam's razor would tend to suggest that it should be rejected in favor of a single-state theory. In the following two chapters I will argue for the plausibility of such a theory. Before moving on to this project I need to address the second of Smith's objections to a *besire* based theory, the conceptuality claim.

Let us re-familiarize ourselves with Smith's conceptuality claim. Smith has claimed that the *besire* theorist must maintain that the belief-like part of the *besire* cannot be separated from the desire-like component, even modally. I read this as a claim about what we can imagine such that, if our moral mental states are *besires* rather than ordinary beliefs coupled with desires then we ought not to be able to imagine the belief-like part existing without the desire-like part or vice versa. If this is correct then the *besire* theorist is indeed in trouble and even a dual-state theory will not save him. The reason for this is that the parts of a *besire* can be pulled apart and conceived of separately. For example, the *besire* referred to a few paragraphs above can be separated, conceptually speaking, into a part that contains the content of the *besire* (it is the holiday season) and the motivational part (I feel that I ought to give to charity).

The fact that *besires* can be thought of as consisting of two parts which can be separated *conceptually* does not undermine the thesis that *besires* actually exist and that the two parts cannot be actually separated. Zangwill makes the following observation: "But if moral beliefs are necessarily motivating, in the way that the motivational internalist has in mind, it is surely *because they are essentially* motivating, just as water is necessarily H<sub>2</sub>O because it is

essentially H<sub>2</sub>O.”<sup>158</sup> A person with very little respect for science can certainly imagine that water is not H<sub>2</sub>O, perhaps because they believe there is no such thing as molecules or atoms. Whatever the cause of the skeptic’s doubts, the truth about the nature of water will remain unchanged. If the nature of some moral mental states is that they really are such that they are structured so that they necessarily motivate then the fact that we can imagine them being otherwise is quite beside the point. Craig DeLancey makes much the same point in the following passage:

The claim that internalism is metaphysically necessary (a kind of necessity that is almost never well defined) is a claim granting to internalism a non-natural status, even though internalism is explicitly concerned with the actual psychology of human beings as they act in the natural world.<sup>159</sup>

Smith might insist that Zangwill and DeLancey miss the point. While we can imagine that H<sub>2</sub>O is not water we can also imagine that it is water. However, we cannot imagine that the belief-like part of a moral mental state is inseparable from its motivational state and this inability to imagine this shows that it is impossible for the two to be inseparable. This objection is a poor one for two reasons. First, it does not seem likely that we cannot imagine the two as being inseparable. In fact, the close connection between moral facts and moral motivation seems to provide significant motivation for various non-cognitivist views of morality. It is doubtful that such views could establish any initial plausibility if it were unimaginable that moral mental states motivated necessarily. Second, what is imaginable does not provide a sure guide to how things actually are. For example, radioactive decay appears to be a probabilistic process all the way down. Scientists can establish half-lives for various substances but whether any particular atom will decay at any particular point in time is purely a matter of chance. Speaking for myself, I find this hard to imagine. It should be the case that for any particular atom we ought to be able to predict, at least in theory if not in practice, that it is closer to or further from decaying, perhaps

---

<sup>158</sup> Zangwill 2008a, p. 52.

<sup>159</sup> DeLancey 2002, p. 146. While DeLancey makes this point during a criticism of internalism he is careful to say that the position he is rejecting is what he calls ‘simple internalism’, the view that *all* moral judgments are intrinsically motivating.

because of structural changes within the atom. Yet just because this is easier for me to imagine does not mean that this is the way things really are. Our beliefs about the nature of radioactive decay should be determined by what our best theories tell us about the nature of radioactive decay.

### 3.6 Interlude on the nature of mental states and our justification for believing in them

Before moving on to the positive argument in favor of the dual-state view that I favor, it would be appropriate to discuss the justification that exists for believing in the existence of mental states. It is not my intention in this section to make an original contribution to the debate concerning whether or not we are justified in believing that there are beliefs and desires. My concern will be the source of such justification as we have and what this justification means for the debate concerning desires. I will argue that whether or not we are justified in believing in the existence of desires comes from their inclusion in or exclusion from our best moral theory.

One possible source for justification of our belief in a particular mental state—and by induction a type of mental state—is introspection. For example, it may be that I am justified in believing that pains exist, at least in my own case, because I know through introspection when I am in pain. I may infer that pains exist in others as well because they exhibit similar kinds of behavior as I do when I am in pain, including saying that they are in pain. My knowledge of other peoples' pains may be due to a theory about similar behaviors having similar causes but, so the story goes, I can know that I am in pain simply through introspection. Call this kind of justification of mental states phenomenological justification.

Another way in which a belief in a mental state can be justified is because the existence of that mental state can be justified through its inclusion in our best theory. At first glance this strikes many people as being unnecessary. Isn't it just obvious that we have beliefs, desires, wishes, fears as well as other mental states? However, Wilfred Sellars argues that many mental states may not be available through introspection in the way that pain is. Wilfred Sellars'

influential thought experiment concerning the proto-human Ryleans given in his essay, “Empiricism and the Philosophy of Mind” is especially relevant.<sup>160</sup> Sellars describes a group of people, the Ryleans, who possess a public language but lack any concepts of inner mental entities such as beliefs, desires, and so on. One day an intelligent Rylean named Jones wonders why it is that his fellow Ryleans can behave intelligently even when they are not speaking. Jones comes to suspect that his fellow Ryleans are talking to themselves, that is, silently, while they are behaving intelligently. Jones teaches this theory to other Ryleans and eventually it becomes common to explain other peoples’ behaviors as well as our own by citing mental states rather than through behavioristic terms. Note that it is compatible with some mental states being known through introspection that not all are.

As I understand Sellars he is suggesting that our belief in (e.g.) beliefs and desires is something that we originally learn because we live in a community of people who make reference to beliefs and desires and we come to speak and think as if they do but we do not learn about beliefs and desires through introspection. If Sellars is right this opens up the possibility that our theory is wrong and that its postulates do not exist. Call this kind of justification of mental states theoretical justification.

Paul Churchland (1981) argues that folk psychology constitutes a theory and a poor one at that.<sup>161</sup> By folk psychology, also sometimes called common-sense psychology, Churchland is referring to our use of propositional attitudes such as beliefs, desires, hopes, and fears to explain and predict behavior. For example, we may predict that Mark will call his mother in the near future because he believes that his mother enjoys hearing from him and he desires that she be happy. We can also use folk psychology to explain behavior. Upon hearing that Mark has called his mother we may explain such behavior by citing a desire to please his mother coupled with a belief that calling her would please her. Churchland argues that folk psychology

---

<sup>160</sup> Sellars 1963.

<sup>161</sup> Churchland 1981.

is a theory whose domain has become much more limited than it used to be (just a few thousand years ago it was used to explain the behaviors of inanimate objects), that it does not provide deep explanations of the causes of behavior and does not fit in well with other scientific theories. Churchland's rejection of folk psychology has not been widely accepted in philosophy. Jerry Fodor has argued that regardless of whether folk psychology is dispensable in theory it is not in actuality and that the theory is both deep and powerful.<sup>162</sup> What seems to be agreed upon, by and large, is that folk psychology is a theory and that to the extent that we are justified in believing in a type of mental state will come from its inclusion in our best theory.

Smith presents his own theory concerning the existence of desires. He starts by arguing against two kinds of phenomenological accounts of desires. The first he calls the strong phenomenological conception of desires.<sup>163</sup> According to this account, desires are simply states with phenomenological content somewhat like sensations. The first problem with this account that Smith discusses is that having a desire and believing that we have a desire can come apart. For example, we can believe that we have a desire that we do not have and we can have a desire that we believe that we do not have. Arguably this is not the case for other phenomenological states such as pain. It is plausible that we believe that we are in pain if and only if we are in pain.

Smith illustrates his point about desires through the use of two examples. In the first, a man regularly goes out of his way to visit a newsstand to buy a newspaper even though there are other places to buy a paper that are more convenient. The main difference between the newsstand he visits and the ones that he does not is that the former has mirrors in which one can hardly avoid looking at oneself. If the man were asked if he visited the newsstand that he does because he could look at himself in the mirror he would sincerely deny it. Yet suppose that, were the mirrors to be removed, he would stop going out of his way to visit that newsstand.

---

<sup>162</sup> Fodor 1987.

<sup>163</sup> Smith 1994, section 4.5, p. 105.



In the circumstances described, it is reasonable to ascribe to the man a desire to look at himself in the mirror even though he does not believe that he possesses such a desire. In his second example, Smith describes a man who has a strong desire to become a great musician. His mother also desires that he become a great musician and the man desires to please his mother. If asked if his desire to become a great musician is in order to please his mother the man would sincerely deny it. However, when the man's mother dies he quickly gives up his pursuit of a career in music for something else. In such a case, Smith says, it would be reasonable to say that the man believed he possessed a desire to become a great musician that he did not really possess.

Another problem for the strong phenomenological conception of desires that Smith points out is that it does not account for the fact that desires have propositional content. We would say of a desire that it is a desire that \_\_\_\_\_ where we would insert in the blank a proposition of some kind. For example, John desires that he learn how to ride a bike, Sally desires that Stephanie calls her mother, Ralph desires that he eat a sandwich, etc. We do not say the same thing about pains or tickles or itches, though they may cause desires. (Theresa may desire that she scratch her leg *because* it itches where the 'because' is read as causal.)

An alternative theory that Smith considers is what I will call the weak phenomenological conception of desires. On this conception, desires possess propositional content but also consist of a phenomenological aspect. This theory is also problematic. Smith's newspaper example discussed above seems to count against this theory as well. The man possesses a desire to look at himself in the mirror but does not realize it. Arguably the desire lacks a phenomenological aspect or else he would at least become aware of it when it was brought to his attention.<sup>164</sup> However, Smith offers an additional example. We frequently attribute desires to people when they appear to have a long term interest in something, like a father who has a

---

<sup>164</sup> Smith's second example is more problematic. The man is aware that he has a desire to become a great musician and a desire to please his mother but is unaware that his desire to become a great musician is an instrumental desire (i.e. that he can please his mother by becoming a great musician).

long term interest in helping his son, even though it is implausible to think that the father is continuously *feeling* an urge to help his son. The desire is felt at various points in time though we plausibly attribute the possession of the appropriate desire to the father even when it is not be felt.

Smith rejects phenomenological conceptions of desires in favor of a dispositional account. Smith discusses a number of advantages of a dispositional theory. First, a dispositional account of desires can explain why we are sometimes unaware of our desires. If a desire is a disposition to behave in certain ways under certain circumstances this does not entail that we are aware of this. For example, a man may have a disposition to defer to the opinions of women but be unaware that he has any such disposition until it is brought to his attention. Second, a dispositional account can provide an explanation of the phenomenological evidence. Whereas some desires lack phenomenological content it may be that others have it and do so necessarily. The reason for this, Smith says, is that the disposition will be a disposition to cause, among other things, the phenomenological content. Third, a dispositional account can explain the extent to which beliefs are an aspect of our desires. Typically, for any desire, if you have a desire to £ then you believe that not £. Finally, a dispositional account can explain the fact that it makes sense, roughly speaking, to think of desires as states that have a direction of fit such that they lead us to try to make the world be a certain way. Since a disposition is a linking of behaviors and circumstances it makes sense to think that we are trying to make the world be a certain way.

### 3.7 Conclusion and preview of the next chapter

In this chapter I have established that it is possible that besires exist and that our justification for believing that they do exist will depend on whether they fit in our best moral theory. It is the task of the next two chapters to try to establish that we have good reason to think that besires will be included as a part of our best moral theory. The justification of the

existence of besires will be more challenging than justifying the existence of desires. The existence of desires is taken for granted by most people. The project then becomes explaining the nature of desires. On the one hand, it seems that we can readily do without the concept of besires so why believe that such mental states exist? On the other hand, it is common to think of there being an especially close connection between a person's accepting a moral fact and that person's being motivated to act on that fact. Besires could play a role in helping to explain this observation.

My case for besires will proceed in two stages. In the next chapter I will argue that besires fit nicely into Aristotelian moral theory and that his theory helps to explain why some of our moral mental states would be likely to be besires. In chapter five I look to the recent work of Ruth Millikan, Tamar Gendler and others to provide a more modern justification for the existence of besires in addition to the more modern notion of moral beliefs and desires.

## Chapter Four

### 4.1. Introduction

In this chapter I argue for a dual-state interpretation of Aristotle's virtue ethics. I begin by reexamining the distinction between internalism and externalism by asking what externalist moral motivation will look like if it exists. With a working definition of externalist motivation in hand, I proceed to an examination of Aristotle's ethical theory. I argue that when we keep in mind that internalism is not a normative theory and that externalism is not a universal theory we will see that Aristotle's theory supports a dual-state position. However, even if the argument fails, we will see that there is an independent argument that demonstrates the presence of externalist motivation. In addition, Aristotle's account of moral education provides a framework through which we can understand why we should expect there to be both internal and external motivation. I end the chapter with a summary as well as a preview of the next chapter.

### 4.2. Internalism, externalism and dual-state motivation

I begin by stipulating a generic definition of motivational moral internalism.

(x) If X makes a moral judgment that A is right then X is necessarily motivated to do A.

The definition does not say anything about whether X actually does A. Some have held that X's making a moral judgment means that X possesses an overriding desire to do A.<sup>165</sup> I will be concerned with a weaker form of motivational internalism in which the agent may fail to do A for various reasons. For example, X may possess a stronger motivation that conflicts with doing A. According to this version of motivational internalism, any motivation no matter how slight qualifies. The key point is that there is a necessary connection between making a moral judgment and being motivated to act in accord with that judgment. This necessary connection can take one of two forms.

---

<sup>165</sup> See (e.g.) Hare 1981 chapter three.

First, the moral judgment may be such that the judgment itself includes motivation. This is how John McDowell thinks of the connection between moral judgment and moral motivation.<sup>166</sup> There is no need to posit an independent desire to act in accord with the moral judgment to explain how moral judgments can motivate.<sup>167</sup> Though it is common to cite an independently existing desire to explain purposive action this ascription is “simply consequential.”<sup>168</sup> We frequently attribute to someone who acts on reason a desire to do as she or he did but this in no way conclusively proves that a desire was present as an additional component. Judging that a situation calls for courage might be sufficient to motivate the virtuous person to behave courageously. This is the kind of motivation that exists in besires if any such mental states exist.

The second kind of moral internalism construes the connection between judgment and motivation as necessary but as involving a belief and desire which are distinct. That is, the judgment is taken as a belief which is distinct from the motivation which is understood to be a desire. This kind of view is defended by Michael Smith.<sup>169</sup> Smith’s view is that our moral judgments express beliefs about objective matters of fact. What we have normative reason<sup>170</sup> to do is a reflection of what we would desire to do if we were fully rational. Second, we can explain the fact that if we have a moral belief then, *ceteris paribus*, we will be motivated to act in accordance with such a belief. This is because, on Smith’s account, beliefs about what our fully rational selves would desire can actually create desires. Yet, on this view, desires and beliefs remain separate existences.

---

<sup>166</sup> See McDowell 1978, 1979 and 1995 among others.

<sup>167</sup> This also seems to be the idea behind what is sometimes referred to as a besire. On besires see Altham 1988, Smith 1994, and Zangwill 2008, though Zangwill appears to conflate the two kinds of moral internalism.

<sup>168</sup> McDowell 1978, p. 15.

<sup>169</sup> Smith 1994.

<sup>170</sup> On Smith’s view, our moral reasons are a proper subset of our normative reasons.

If the views of both McDowell and Smith count as forms of internalism it is worth asking what distinguishes externalism from internalism. Externalism is sometimes characterized simply as the denial of internalism.<sup>171</sup>

If X is a moral agent and X makes a moral judgment that A is right then it is possible that X is not motivated to do A.

Three things should be noted here about this definition of moral externalism. First, the denial of internalism is compatible with the existence of moral agents who make moral judgments and are always motivated to act in accordance with them as long as there are some moral agents who are not like this. This is important because truly virtuous agents, should any actually exist, may fall into this category. Second, externalism is compatible with the existence of an agent who makes some moral judgments that necessarily motivate and other moral judgments that motivate only contingently. Third, the discussion of internalism and externalism tends to be factual rather than normative; it focuses on whether internalism is true rather than whether it would be better for a moral judgment to motivate internally or externally. The definition of motivational externalism given above fails to distinguish between those who believe that all moral judgments motivate contingently and those who believe that some judgments motivate contingently and that some motivate in the one of the internalist ways mentioned above. Let me then redefine moral externalism and distinguish it from what I am calling a dual-state theory.

Moral externalism: (X) (y) If X is a moral agent and X makes a moral judgment y then it is always a contingent matter that X is motivated to act on y.

Dual-state theory: Some moral judgments are internalist and some are externalist.

What must externalist moral motivation look like if it exists? Though we saw in chapter two that Smith's argument against moral externalism is weak, it will be helpful to consider an objection he raises. He is concerned that moral externalism amounts to what he calls a kind of moral fetishism. Smith tells us that "good people care non-derivatively about honesty . . . their

---

<sup>171</sup> See e.g. Brink 1989, p. 42.

children and friends,” and so on.<sup>172</sup> To underscore the point, Smith reminds us of Bernard Williams’ point that a woman whose husband saved her because *it was the right thing to do* and not because she was his wife would be considered to have *one thought too many*. Smith’s point, I take it, is that if a decision takes the form that the externalist thinks is possible then rather than caring directly about a person or a way of behaving (e.g. honestly) the person will help a friend or behave honestly because it is the right thing to do.

If moral motivation for any particular judgment is external to the judgment then it is a contingent matter whether the judgment motivates. The idea might be that there is *a pool of motivation* that exists separately from moral judgment and which is not automatically activated when a person makes a moral judgment. How exactly this pool of motivation will be activated in particular cases is a difficult question but it must involve some recognition on the part of the agent that she or he is in a moral situation.<sup>173</sup> By a moral situation I mean a situation in which morality demands a certain kind of response that the agent was not necessarily going to make anyway. In some sense almost every situation is a moral situation. For example, the situation I am in right now demands that I not stop what I am doing and light the apartment building in which I live on fire. But I wasn’t going to do that anyway. The kinds of situations to which I am referring are ones in which the morally correct responses are not ones that we were going to do anyway.

It is important to note that the externalist is not denying that there is a reliable connection between making a moral judgment and being motivated accordingly. Brooke Sadler characterizes externalist moral motivation as a motivation “to do what is right, whatever that is.”<sup>174</sup> This seems to be the way the externalist needs to characterize moral motivation. For suppose that the attachment of motivation to moral judgments was a purely hit and miss matter. If this is the case then the recognition of an action with moral significance is no more likely to

---

<sup>172</sup> Smith p. 75.

<sup>173</sup> This “pool of motivation” metaphor will be discussed in more detail in section five.

<sup>174</sup> Sadler 2003, p. 68.

provide motivation than any other possible action that can be undertaken. Imagine I recognize that I could get up and turn on the radio and that doing so or refraining from doing so has no significant moral relevance. I may be motivated to do so and I may not. Perhaps I prefer to continue watching television. On the other hand, suppose I recognize that calling my mother on her birthday (today) would make her happy and that this would be the right thing to do. If externalism about moral motivation is like any other motivation then I would be no more likely to call my mother than not. If I like talking to my mother I'll probably call her but if I am busy I may not get around to it. People would be amoral, at least when they were not being motivated internally. While externalists want to allow for the possibility of amoralism, they do not want it to be the norm for agents. Therefore, it is more plausible to think that there is a general pool of moral motivation. Can we make sense of such a characterization within an Aristotelian framework?

#### 4.3 The role of reason and desire in virtuous and morally continent persons

Aristotle identifies two positive moral states that are possible for human beings, virtue and moral continence.<sup>175</sup> Put in a situation in which (e.g.) generous behavior is called for, both the virtuous and the morally continent person will do the right thing but they do so in different ways, as the following passage illustrates.

No doubt, however, we must none the less suppose that in the soul too there is something contrary to the rational principle, resisting and opposing it. In what sense it is distinct from the other elements does not concern us. Now even this seems to have a share in a rational principle, as we said; at any rate in the continent man it obeys the rational principle—and presumably in the temperate and brave man it is still more obedient; for in him it speaks, on all matters, with the same voice as the rational principle.<sup>176</sup>

---

<sup>175</sup> *Nicomachean Ethics* 1145a 15-20. Aristotle also identifies a superhuman kind of virtue that is more suited to gods than human beings but has little to say about it.

<sup>176</sup> *Nicomachean Ethics* 1102b 24-29.



Particularly important for understanding this difference are two parts of the human soul, reason and desire.<sup>177</sup> Desire is that part of the soul that can resist and oppose the rational part. In the incontinent person desire does not obey reason as it does in continent and virtuous persons. The temperate and brave man (person) Aristotle refers to will be someone who is virtuous, as Aristotle holds that someone who possesses even one of the virtues will possess all of them.

For Aristotle, reason is the part of the soul that makes us distinctively human and, because the virtuous person will be one who is functioning correctly, reason will play an important role.<sup>178</sup> A necessary condition for being a virtuous agent is that the reasoning part and the desiring part of the soul are related in the proper way. To understand how the relation between desire and reason in the morally continent and the virtuous person differs, it is helpful to think of desire as a dog and reason as the dog's owner. In the virtuous person, desire is like a well-trained dog that does what its owner commands without any internal conflict. The owner says "sit" so the dog sits. The squirrel running through the yard becomes irrelevant to the dog's behavior even though, had the dog not received a command from its owner, the dog would have chased the squirrel. In the continent person, desire is like a dog that has not been properly trained but has received some training and can respond to that training. The result of this inadequate training is internal conflict. On the one hand, the dictates of reason carry weight but on the other hand, there is a desire to do something else. In the continent person, the dictates of reason overpower desire and the agent follows reason. Returning to the dog example, the dog overcomes the powerful urge to chase the squirrel. Though the end result is the same for the continent person and the virtuous person, the presence of an internal conflict opens up the possibility that desire may win-out and the person may do that which they know is wrong.

Aristotle's account of the interaction between desire and reason is somewhat similar to that of David Hume. They agree that only desire can motivate action. Hume states that "reason

---

<sup>177</sup> In discussing Aristotle I will sometimes speak as if reason and desire are distinct parts and this is often the way he writes. However, reason and desire are better thought of as different capacities.

<sup>178</sup> *Nicomachean Ethics* Book 1, chapter 7.

alone can never be a motive to any action of the will and . . . that it can never oppose passion in the direction of the will.”<sup>179</sup> But they disagree on what it would be to call a desire unreasonable. For Hume, there are only two situations in which it makes sense to call a desire unreasonable. First, when we have desires about things that don’t exist. If I desire to protect myself from ghosts this counts as an unreasonable desire because ghosts don’t exist. Second, an instrumental desire may be deemed unreasonable when it provides an insufficient means to fulfilling another desire a person has. For example, suppose that I desire to make as much money as possible and, in pursuit of that end, I desire to quit my full time job at which I am paid \$15 an hour to take a job at which I will be paid a salary of \$25,000 a year. My desire to quit my current job and accept the other job is unreasonable because earning \$ 31,200 a year, which I can expect to do at my current job, is more than earning \$25,000 at the new job.<sup>180</sup> Aristotle has a much more robust view of what counts as an unreasonable desire. For Aristotle, it makes no sense to say that a desire uninformed by reason is reasonable. Although the presence of desire is a necessary condition for setting an action into motion, ideally it is reason that sets the goals and not desire.

According to Aristotle, the ideal moral state for human beings is virtue and when a person acts on virtue they are clearly internally motivated. Aristotle says that in the virtuous person reason and desire “speak in the same voice.” However, internalism is not a claim that covers only the best moral states such that it is best that when a person makes a moral judgment that person is motivated to act in accord with that judgment in one of the two ways mentioned above. Internalism is a claim that covers all moral judgments. This may appear to open up the possibility that in the case of the continent or incontinent person, when there is motivation accompanying a moral judgment, it will be external in the way specified in section one.

---

<sup>179</sup> Hume 1978, p. 413.

<sup>180</sup> Ibid, p. 416.

I believe that this is a mistaken interpretation of Aristotle. The main distinction between the virtuous person and the continent one concerns the presence of contrary desires; the latter has at least one whereas the virtuous person has none. The presence of contrary desires does not indicate that the continent person is not motivated by her judgment that a particular action is the right thing to do. In fact, the conflict seems to arise from the failure of the judgment to be absolutely authoritative rather than the judgment not being accompanied by moral motivation. The source of externalist motivation in Aristotle will need to come from a somewhat different source.

#### 4.4 Desire and reason

Toward the end of *De Anima*, Aristotle considers the problem of how a person's soul produces bodily movement. With characteristic reasonability, he begins by noting that both reason and desire play a role in producing movement, desire starts the movement and reason provides the means to accomplish the end that desire has set. Though the two work together, it is desire itself that is the true internal originator of movement. "As it is, mind is never found producing movement without appetite . . . but appetite can originate movement contrary to calculation."<sup>181</sup> However, desire is directed toward some good which is what moves desire but is itself unmoved. "That which moves without itself being moved is the realizable good."<sup>182</sup> So the object of our desire is the unmoved mover which then causes movement in the appetitive part of the soul. Reason may play a role in this movement but it is appetite that is the part of the individual that causes movement. Given that Aristotle conceives of desire and reason as being separate parts of the soul (or separate faculties) which are related to each other in different ways in virtuous persons than they are in morally continent ones and given that desire is seen as being obedient to reason in the virtuous person, a strict reading of Aristotle seems to provide

---

<sup>181</sup> *De Anima* 433a 23-25.

<sup>182</sup> *De Anima* 433b 15-16.

more support for the second kind of internalism as being the state that exists in the virtuous person than the first.<sup>183</sup>

Henry S. Richardson argues that the faculty of desire in reasoning and non-reasoning animals is aimed at the good.<sup>184</sup> It is the (apparent) good itself that is the originator of movement and which causes movement in the faculty of desire which then causes movement in the animal. For human beings (i.e. rational animals), mind may also be involved. On the distinction between human and animal capacities for the good, Richardson says the following:

Human deliberative capacities mark a departure (between the way Aristotle would account for human action and animal movement) not because they introduce a new relation to the good, but because they involve abilities to make explicit comparisons and to follow inferences in a way that allows agents to deal rationally with conflicts among different aspects of their good.<sup>185</sup>

This seems consistent with Aristotle's teleological approach to the behavior of living beings.

The soul of a living being aims at the good for that type of being. Non-rational animals lack a reasoning part of their soul but their behavior is still aimed at what is good for them. Desire is directed at the general good though it may not always hit the mark. Desire sees the immediate consequences of an action only; reason is needed for judging long-term consequences.<sup>186</sup>

Sometimes the two will conflict. For example, pleasure is a good but taking pleasure in the wrong things, such as the enjoyment derived from stolen goods or drinking one too many glasses of wine, is bad. According to Aristotle, the ideal life for human beings is the eudaimon life and reason is a necessary component to the achievement of such a life. Given Aristotle's teleological conception of the soul and the actions produced by it coupled with the ability of desire to be responsive to reason, it looks like Aristotle's theory supports the idea that there is a pool of moral motivation (i.e. motivation aimed at the good) which exists independently of the judgments produced by reason. To be clear, there is a sense in which Aristotle's account is

---

<sup>183</sup> See the distinction made between the two kinds of internalism at the beginning of this chapter.

<sup>184</sup> Richardson (1992) p. 382.

<sup>185</sup> Richardson, p. 399.

<sup>186</sup> *De Anima* 433b 4-11.

internalist. We are always motivated to pursue desire's judgments about the good. Reason makes pronouncements on the same subject but whether these will be accompanied by desire seems to be a contingent matter, at least in some agents. Yet internalism seems to be an account of the pronouncements of reason.

#### 4.5 Objection 1 The lack of phronesis in the morally (in)continent

Why has Aristotle been assumed to be an internalist even though reason and desire are distinct? Kristjan Kristjánsson argues for an Aristotelian form of motivational externalism but stops short of judging Aristotle himself to be an externalist.<sup>187</sup> He raises the concern that a person who is not fully virtuous cannot possess practical wisdom (phronesis) and vice versa. As a result, the person who possesses only external motivation will fail to be practically rational. Kristjánsson rejects an interpretation by Brad Inwood which recognizes "comprehension (sunesis) as an intellectual virtue with the same subject matter as phronesis."<sup>188</sup> Kristjánsson rejects this reading because he believes that sunesis is concerned primarily with what other people do rather than through guiding our own action.

It is not clear that this is the correct interpretation. Aristotle's recognition of varying levels of continence in addition to full virtue suggests a different interpretation. The person who comprehends that he ought to do X will not be moved in the same way as would the practically wise person who recognizes that he ought to do X. Aristotle recognized moral weakness as a problem and one that struck counter to intuition. How can we make a judgment about what it is best to do and yet fail to be moved by that judgment? The judgment comes from within the individual and the internal moving force is within us as well. Sometimes though, we fail to be moved. It is not mysterious, however, if a person judges of another that she ought to do X but the judger herself remains unmoved. The resulting judgment does not apply to what should

---

<sup>187</sup> Kristjánsson 2012.

<sup>188</sup> Kristjánsson, p. 19.

move her. In addition, the judger's judgment cannot move the other person because the motivating force is not within the one being judged. In that sense, when I comprehend that I ought to do X but, oddly enough, remain unmotivated it is as if I am making a judgment about someone else, which would explain the lack of motivation. This may be what Aristotle meant.

Even if this interpretation of Aristotle's meaning is misguided, it is still clear that Aristotle believes that the incontinent person's failure to do what is right does not stem from a lack of possession of knowledge concerning the right action to be performed in the particular circumstance. In some sense, the incontinent person knows what is right.<sup>189</sup> Aristotle considers whether the incontinent person may possess the universal premise of the practical syllogism but not the particular premise and fails to get to the conclusion (and fails to act). Yet clearly an incontinent person can get to the conclusion, "I ought to do X in the circumstances I am in right here and now" and yet fail to be motivated. Just what the incontinent person is lacking is phronesis and it is not clear that phronesis is not primarily having one's motivations and judgments interconnected in the way that they are in the virtuous person. If this is correct then citing the lack of phronesis in the incontinent person cannot be used to debunk externalism as it begs the question against the moral externalist. It is important to remember that internalism is not a normative theory.

#### 4.6 Objection 2 The directedness of desire

Aristotle has identified two different functions in the soul, reason and desire, which have somewhat different roles. They both have a role in producing action but only desire can produce action on its own. Aristotle's notion of desire is more complex than some other accounts of desire. Dennis Stampe argues for an Aristotelian-style account of desires. He maintains that desires are representations of the apparent good.<sup>190</sup> On Stampe's view, desire

---

<sup>189</sup> *Nicomachean Ethics*, Book 7, chapter 3.

<sup>190</sup> See Stampe 1987 and 1990.

involves the perception of bodily states. As a result, reasoning from desire starts from a perception, proceeds to an instrumental observation about how one could bring about what is perceived as good, and concludes with an intention. On Stampe's view, desires have both representational and motivational aspects. Their structure is essentially that of a desire. I will not critically examine the implications of Stampe's view here though I intend to return to it in future work.

Returning to Aristotle's account, reason can influence desire because desire listens to reason. In the virtuous person, desire and reason act as one with reason setting the goal. In the (in)continent person, desire listens to reason but whether reason will guide the action depends upon how desire resolves the conflict between the goal it has given itself and the goal given to it by reason. A way of thinking of this that is consistent with Aristotle's teleological approach is to see reason as having the ability to create a goal within desire.

This interpretation of what is going on in the (in)continent person is broadly compatible with Smith's theory in which a normative moral reason gives rise to a corresponding desire. In some cases, the new goal which is created within desire will then need to compete with other goals with which it may conflict. (Aristotle is not clear on how these conflicts are resolved.) Desire contains both general appetites and specific, acquired desires. Sometimes these desires allow us to satisfy an appetite, as my specific desire for Pad Thai allows me to satisfy my hunger in a particularly pleasurable way. Reason can also create other kinds of desires as well. The recognition that one is in a situation that calls for one to be resolute in the face of danger, for example, can lead to a desire to be resolute in the face of danger.

Such an account may seem to leave no room for externalist moral motivation. Desire contains the general appetites in addition to acquired desires which either satisfy the general appetites or which are dictated by Reason. Where is the "general pool of motivation" *to do what is right whatever that happens to be* that is necessary for externalist motivation? If such a pool does not exist then a dual-state theory will be ruled out as well. Yet the existence of such a

“pool” is exactly what we should expect if the relationship between reason and desire is as it has been portrayed above. What I will mean by a pool of motivation is an acquired desire to do what is right what that turns out to be. To explain why we should expect this I will need to revisit chapter two’s discussion of *de dicto* and *de re* desires and their relevance to Smith’s claim that moral externalism requires a kind of motivational fetishism. We saw in chapter two that there were good reasons to think that if there were *de dicto* motivation then it would not be fetishistic.<sup>191</sup> Here I will argue that there is *de dicto* motivation.

To show that there is *de dicto* moral motivation I will revisit some of the cases discussed in chapter two. It is important for the externalist to establish the existence of cases in which externalist motivation is needed. The reason is that even though the presence of *de dicto* motivation in these cases is not fetishistic the internalist has explanations available for these cases. There are two types of cases which occur frequently in the literature which purport to show that the presence of *de dicto* desire would not be fetishistic, at least under some circumstances. The first case I will call *temptation*. The example below is from Lillehammer.

Consider someone who goes to a party during a phase when she is tired of her husband. At the party she meets a very charming person and is tempted to have an affair. She judges that it would be wrong to have an affair on account of her husband’s feelings. But she is temporarily indifferent to her husband’s feelings. However, she has a standing *de dicto* desire to do what is right which, together with her moral judgment, causes her to do the right thing, in spite of the absence of a *de re* desire to do the right thing and the presence of a *de re* desire to do the wrong thing.<sup>192</sup>

Do we need to appeal to *de dicto* motivation to explain the woman’s remaining faithful to her husband? The woman has two desires which cannot both be satisfied. That is, she cannot remain faithful to her husband and have an affair. If she were not temporarily indifferent to her husband’s feelings, perhaps because they have been arguing recently, she may not even be

---

<sup>191</sup> See section 2.5.

<sup>192</sup> Lillehammer, Hallvard. 1997. “Smith on Moral Fetishism.” *Analysis*, 57.3, 187-195. See Sadler, Brooke. 2003. “The Possibility of Amoralism: A Defense Against Internalism.” *Philosophy*, Vol. 78, No. 303, 63-78 and Olson, Jonas. 2002. “Are Desires *De Dicto* Fetishistic?” *Inquiry: An Interdisciplinary Journal of Philosophy*, 45:1, 89-96 for similar examples.



tempted by the charming stranger. Admittedly, possessing a de dicto desire to do what is right could help the woman avoid cheating on her husband.

Yet, we need not posit a de dicto desire to explain why the woman refrains from cheating on her husband. We could recommend to the woman that she attend to the right-making features of the situation.<sup>193</sup> The more she attends to these features (i.e. the reasons she has to not have an affair), the more likely she is to act rationally and not cheat on her husband. Further, this solution respects the phenomenology of cases of moral temptation. Consider what the woman would do in the situation described in order to try to resist the temptation to cheat. She would probably recite the reasons why she should not have an affair (e.g. 'He is my husband', 'I would be deeply hurt if he did this to me', 'It will ruin our relationship', 'I took an oath', etc.). Suppose reciting these reasons is effective and the woman resists temptation. The internalist can plausibly say that focusing on the reasons to be faithful is what allowed the woman to resist temptation, either by strengthening her desire to remain faithful or dampening her desire to cheat or both.

The second case I will call the case of moral uncertainty.<sup>194</sup> Suppose that a person (Sol) becomes aware of the possibility of warning a fellow co-worker (Sally) that she may potentially be the victim of gender-based discrimination in the workplace. They have a male co-worker whom Sol has good reason to believe is quite sexist and may be in position to undermine the woman. Sol forms a desire to warn Sally. However, Sol also knows that his sexist co-worker is in tight with the boss and that warning Sally could put Sol's own career at risk. He is also not sure about how likely Sally is to be discriminated against or how much being warned about the risk would benefit her. Sol discerns different morally relevant factors, some of which indicate

---

<sup>193</sup> Toppin, Teemu. 2004. "Moral Fetishism Revisited." *Proceedings of the Aristotelian Society*, 104, (3), 305-313, makes a similar point.

<sup>194</sup> For discussions of moral uncertainty see Dreier, James. 2000. "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research*, Vol. 61, No. 3, 619-638, and Sadler 2003.

that he ought to put himself at risk and warn Sally and others which indicate that he should refrain from doing so. It is not clear what he ought to do.

The possession of de dicto moral motivation would be (morally) useful to Sol in this situation. His uncertainty about what to do prevents him from acting immediately but having de dicto motivation could lead him to do be sensitive to characteristics of the situation which he may become aware of in the future that may reveal what is the right thing to do. Yet, we need not appeal to de dicto motivation in situations of moral uncertainty either. In fact, the right-making features are present in the situation described. What they show is that Sol should form a desire to seek out more information about the situation. If the additional information indicates that it would be right to warn his co-worker he can form a desire at that time to do so. Neither in the present nor at the later point does Sol need to possess a de dicto desire to be properly motivated. Below I give a case that does require de dicto motivation.

Let's start by considering the following (arguably) non-moral example. Suppose that I have a brother who I have learned is leaning toward making an unwise life choice. Specifically, his longtime girlfriend has been pressuring him to get married but he refuses to propose. I have good reason to believe that my brother is quite happy in the relationship and that it would be good for him to get married. I also have good reason to believe that his girlfriend is preparing to leave him if he does not propose relatively soon. I form a de re desire to talk to my brother in an effort to persuade him to propose to his girlfriend. The desire is de re because the desire forms as a result of my recognition of certain right making features in the environment, such as that my brother is jeopardizing his relationship with his girlfriend, my brother will be lonely without her, and so on.

Now let's expand the example. Unfortunately, my brother is stubborn. Following the conversation, I realize that my discussion with him probably has not changed his mind. But my brother is *important* to me and I form a new desire to return to the topic again when the time seems right. This new desire can still be understood as a de re desire. It is a desire to monitor

my interactions with my brother to look for features of future situations, such as that he is in a good mood and a sufficient amount of time has passed, that indicate that it would be the right time to talk to him. Once these features are present I can act once more on my original *de re* desire to convince my brother that he ought to propose to his girlfriend. So far, the example is analogous to the case of Sol and his co-workers in that an original *de re* desire is formed and a second *de re* desire is formed because it is necessary to help satisfy the original desire.

Now let's expand the example even further. Let's say that my brother's decision to not marry his girlfriend comes as no surprise to me as he frequently makes bad decisions in life and this is just the latest of a number of bad choices on his part. As a loving brother, I want to provide good advice. Further, I can reasonably conclude that in addition to the issue with his girlfriend there are other bad decisions that my brother will contemplate making in the future. This kind of situation is likely to *recur* and I know this through my experience with my brother. However, there is not a specific future bad decision of which I can currently be aware. He may get involved in a pyramid scheme or decide to quit his job and try to make a living hustling pool. In a case like this, it is plausible that I will form a general, *de dicto* desire to attend to the important aspects of my brother's life. That is, I will be looking for the presence of features that indicate that it would be right for me to intervene in my brother's life to prevent him from making some serious mistake. This desire must be *de dicto* because there is not some specific action that is its object. It is a general desire to do what I can to help my brother's life go well.

There appear to be three characteristics which, when present in a person, will lead to *de dicto* desire formation in that person. I argue by *reductio*. First, how things go within the domain must be important to the person. Suppose that how my brother's life goes does not matter to me and that I have had nothing to do with him for years. In such a case, we would not expect that I would have any general desire to see his life go well. Second, there needs to be the possibility that there will be non-ideal future outcomes in the domain in question and that the agent believes that she or he has some power over how things go in that domain. Non-ideal

here should be read as relative to the agent who would form the desire. If all outcomes within the domain will be ideal then there is no need to form a desire because all will be as the agent would desire. Further, if the agent has no power over how things go in that domain the best thing the agent can do is wish for things to go better. If my brother always makes good choices in life or if I have no influence whatsoever over him then there will be no need for a de dicto desire because such a desire would serve no purpose. Third, the person must be in a non-ideal epistemic position, either because he is flawed or the situation is one in which certain important pieces of information are obscured, or both. Suppose that I will always become immediately aware anytime my brother contemplates making a bad decision. Let us suppose that both of us have electronic devices implanted in our brains and that whenever my brother contemplates a bad decision this information is automatically transmitted into my brain. Under such circumstances, there will be no need for me to form a de dicto desire to monitor the situation because I will always immediately become aware of circumstances in which I should intervene. As a result, I could count on a de re desire being formed. Of course, this is not the case. I do not always know when my brother is going to make a bad decision. Does the moral realm have the same relevant characteristics?

First, people do place importance on morality and this importance is stressed on many of us at an early age. Moral transgressions are condemned and sometimes punished whereas supererogatory acts are praised and sometimes rewarded. Second, that things will be less ideal than they could within the moral realm I will take as a given. Third, the presence of moral dilemmas, be they genuine or merely apparent, is commonplace. If there are genuine moral dilemmas it will not always be immediately obvious when this is the type of situation a person is in. The good moral agent will be forced to struggle to see if the right making characteristics favor one choice over another before concluding that there is no one right answer. If a dilemma is merely apparent then some work will be required on the part of the good person to determine what is the right thing to do. Yet we need not assume the presence of a dilemma. We

frequently lack information simply because we have not spoken to the right people or been fortunate enough to observe the relevant events. Worse yet, people will often hide relevant information or outright lie about it.

It might be objected that the example I have employed to demonstrate the presence of de dicto motivation presupposes what it is trying to show. By stipulating that my brother is important to me am I not simply assuming the presence of a de dicto desire that my brother's life go well? Fortunately, no such assumption is required. We can posit the importance of a domain to an agent without presupposing de dicto motivation. Suppose that many years ago I became aware that my then eighteen year old brother was going to buy a sports car that he could not afford. I responded to the relevant aspects of the situation by developing a de re desire to dissuade my brother from making the irresponsible purchase. However, I had not yet formed the de dicto desire that my brother's life go well because, prior to this situation I had never really concerned myself with my brother's decision making. That was the responsibility of my parents. Only after having become aware of a number of incidents in which my brother made bad decisions did I develop a de dicto desire. What appears to have been overlooked in past discussions of this topic seems to be that good persons will develop de dicto motivation over time in response to non-ideal conditions. Given the considerations discussed above, I conclude that good persons will acquire this kind of motivation.

I have argued that the presence of external moral motivation in Aristotle's moral philosophy is to be expected because of the manner in which he believes that desire and reason interact within the human soul. Further, Aristotle is best interpreted as a dual-state theorist. Fortunately, Aristotle's account of moral education is more plausible than his account of the soul and it too supports a dual-state theory.

#### 4.7 Aristotle's account of moral education

According to Aristotle, becoming a fully virtuous person is only something that can be achieved well into adulthood. It is helpful if one starts out with natural virtue. In some people, whom Aristotle refers to as “the truly fortunate ones” (1179b 24), there will be a natural tendency toward at least some of the virtues. It is not clear how important the role of natural virtue is to those who eventually acquire full virtue but it will certainly be easier to acquire the habits necessary for full virtue if one is already naturally inclined to behave as the virtues dictate. Nowhere though does Aristotle say that natural virtue is necessary for virtue proper but he does say that “the virtues arise in us neither by nature nor against nature. Rather, we are by nature able to acquire them, and we are completed through habit.” (1103a 24-27).

A second and essential step for developing full virtue is to acquire good habits. In comparing the acquisition of a virtue to the acquisition of a craft, such as carpentry or harp playing, Aristotle says that “we become just by doing just actions, temperate by doing temperate actions, brave by doing brave actions” (1103b 1-2). It is what we learn to do when we confront situations that provoke certain feelings, such as fear of bodily harm or the desire for pleasure, and the feelings that result from being conditioned to do the right thing, that plays a key role in making us virtuous.<sup>195</sup> Habituation plays a number of important roles in helping one to attain full virtue. First, it trains desire to respond in a fully obedient way to reason. Becoming conditioned to respond temperately, for example, requires that other desires, such as the pleasure that would come from having a second piece of chocolate cake, never even enter into our conscious minds as a consideration in favor of what would be overindulging. Second, it prepares us to appreciate that which is fine.<sup>196</sup> Aristotle does not recognize a conflict between what is in our best interest and doing what is virtuous. This is because virtuous actions are pleasant because they are fine and being trained to behave as the fully virtuous person does allows for the formation of the ability to appreciate the pleasure that comes from doing fine actions. This is a

---

<sup>195</sup> NE 1103 b14-22.

<sup>196</sup> In many translations, kalos, is translated as noble rather than fine. I will use fine here.

key reason why moral arguments have limited power to persuade those without a proper upbringing whereas they are more likely to persuade those with a proper moral upbringing. Third, it provides some ability to recognize the morally relevant characteristics of a situation. Nancy Sherman argues that the process of habituation cannot be completely non-cognitive.<sup>197</sup> For example, a person will learn to identify characteristics of a situation that warrant a certain behavioral response.

Habituation takes us a long way toward attaining full virtue but it does not get us all the way there. The third aspect is to listen to arguments so that one who has been properly habituated knows *why* a temperate act is to be chosen and not just *that* it is to be chosen. Whereas the second step works primarily on desire, this third step focuses on reason. The person who has reached the third step is now ready to hear lectures on ethics (or read the *Nicomachean Ethics*). Why is it not okay to stop at the second step? The properly habituated person has learned to appreciate the pleasures that come from doing fine actions but pleasures can be taken, and pains avoided, by doing the wrong things as well. Having a second piece of cake or a third or fourth glass of wine is pleasurable as well. Running away from a battle may help one avoid pain or even death. Why not attain pleasures and avoid pains in these ways instead? By reading the *Nicomachean Ethics* we can learn what courage and temperance are as well as what their corresponding vices are.<sup>198</sup> We can also learn that it is easy to miss the mark and that, for most of us at least, we may be tempted to seek our pleasures or avoid our pains in licentious or cowardly acts, respectively.

We can also come to have a conception of the good life in general (eudaimonia) and how virtuous acts are conducive to such a life whereas vicious ones are not. For example, we may question whether it is best for an individual to care more about her- or himself or whether that individual ought to care for others more than her- or himself. If we become convinced that

---

<sup>197</sup> Sherman 1989.

<sup>198</sup> In addition to other learning about other virtues and vices.

one ought to care most of all for oneself we might be tempted to behave stingily rather than generously. Yet, this conflict between one's behaving virtuously and one's behaving in a way that is in one's own best interest is only apparent. The virtuous person loves the fine and serves his or her own interests by doing what is virtuous.<sup>199</sup>

Though it does not prove that a dual-state theory is correct, Aristotle's theory of moral motivation provides support for such an account. People who receive an ideal moral education will be able to use their intellect to recognize the good and their desire will recognize the intellect's authority on the matter. As a result, they will not desire any apparent goods when these conflict with the judgment of the intellect. From an evolutionary point of view, it probably makes sense for there to be a strong connection between important behavioral norms within our communities. It makes sense that there would be an automatic motivational response in situations that call for adherence to behavioral norms. However, even if fully virtuous people are possible, many people's moral education will be lacking in some respects. Yet, we ought not to think of this as an all-or-nothing situation. As is the case with the incontinent and the continent, we should expect that their intellect will have some ability to recognize the good. But only desire has the ability to initiate action. As a result, if the improperly trained person is to sometimes behave correctly, there will be helpful for there to be a general desire to do as reason dictates rather than as desire originally does. In addition, we may find ourselves in novel situations that call for behaviors that were not covered by our training. Thus, we should not be surprised if there exists a less automatic system that connects situations with motivational responses.<sup>200</sup> As was shown in the previous section, we can expect that moral agents will develop a *de dicto* externalist moral motivation to do what is right whatever that turns out to be. Further, if circumstances dictate that even the virtuous person will not always be able to

---

<sup>199</sup> NE book 9, chapter 8.

<sup>200</sup> Some evidence in social psychology appears to support this interpretation and will be discussed briefly in the next chapter. I intend to return to this work in more detail at a later point.



determine what the right thing to do is then we can expect that he or she will also develop externalist moral motivation.

#### 4.8 Conclusion

If the picture sketched above of Aristotle is correct then every (non-deformed) human being possesses motivation to do *what is right whatever that turns out to be* because we all possess the faculty of desire and desire is aimed at the apparent good. Reason allows us to resolve conflicts between things that appear as good and to take into consideration our long-term interests. However, reason relies on desire for its motivating power. The *Nichomachean Ethics* is largely a blueprint for transforming this beginning state into an ideally functioning rational agent by training Desire so thoroughly that it obediently and willingly follows our moral judgment. The *Nichomachean Ethics* and *De Anima* do not show that there is no such thing as external moral motivation, quite the opposite. The *Nichomachean Ethics* does show that with the proper training we may have no need of it, which is why learning good habits when we are young is so important. Yet moral internalism is neither the claim that only some motivation is internal nor is it the claim that when motivation is internal this is better than when it is external.

Even if this interpretation of Aristotle is misguided there are independent reasons for thinking there is externalist motivation. I have tried to show this section 5 and have also argued in section 6 that Aristotle's account of moral education also supports such a view. In the next chapter we will see that there are additional reasons to believe that motivation is external.

## Chapter Five

### 5.1 Summary of our progress so far and the project for this chapter

The overall project of this dissertation is to provide a naturalistic solution to the moral problem as it is laid out in Michael Smith's, *The Moral Problem*. Addressing the internalism/externalism question as it pertains to motivation in moral philosophy plays an important role in finding such a solution. In chapter one, we surveyed a number of positions that have been taken in the debate and fixed on the most plausible versions of externalism and internalism. Chapters two and three focus on this debate. Chapter two focuses on Smith's rejection of externalism. A number of objections were raised to Smith's argument against externalism. Central to this is the rejection of Smith's claim that the good and strong-willed moral person cannot be motivated by external moral motivation. The third chapter turns to an examination of Smith's favored solution to the moral problem. While Smith provides an interesting structural solution to the moral problem in that it gives us an account of the normative reasons that people have, it does not explain why we should believe that there are normative *moral* reasons. Smith's argument against *besires* was examined in the second half of the third chapter. Smith's argument against *besires* does not go through and a different argument from Nick Zangwill against *besires* was considered. Zangwill's argument only succeeds if we assume a single-state theory in which moral motivation is always external or always internal. Chapters four and five provide a defense of a dual-state theory in which moral motivation is sometimes internal and sometimes external. In chapter four, it was argued that Aristotle's moral philosophy can be interpreted as a form of dual-state theory or that it is at least amenable to such a theory—depending on how we construe Aristotle's view of the interaction between desire and reason. In considering Aristotle's view of the mind we discovered that there was good reason to believe that some motivation would be external. Aristotle's account of moral education and the need to follow social norms make it that much more likely that at least some motivation will be external.

In this chapter I will argue for two main claims. The first claim is that besires actually exist.<sup>201</sup> Part of the defense of this first claim has been addressed in the previous chapter in which it was argued that it seemed plausible that evolution would have selected for a close link between the cognitive recognition of the situation one is in and a behavioral response. The second claim is that the descriptive side of besires corresponds to some objective moral truth. To make a convincing case for these two claims I consider a number of objections to a dual state view. Following a brief review of the moral problem I consider the first objection. First, it might be argued that no sense can be made out of the idea that there is a mental state with aspects of both beliefs and desires, which is what a besire is. In section 5.3, I discuss three modern theories that argue for mental states which have the same kind of structure as besires, Dennis W. Stampe's theory of desire, Tamar Szabo Gendler's theory of aliefs and Ruth Garrett Millikan's theory of pushmi-pullyus. I argue that not only are these theories interesting in their own right but that they receive support from a widely-held theory in psychology. This will complete my defense of the first claim.

Section 5.4 considers an objection put forward by Sharon Street that accounts like the ones discussed in section 5.3 cannot plausibly underlie moral realism. I argue that the realist can respond to Street's view if we do not think of the individual as tracking a distinct moral property, like goodness or badness, through a domain but rather by adopting a version of virtue ethics. In section 5.5 I consider a third objection which is sometimes called the eudaimonism objection. The concern is that even if we have found a plausible candidate for a mental state with both directions of fit, Aristotelian virtue ethics cannot provide an objective basis for moral truths. I argue that a proper understanding of Aristotle's project in the *Nicomachean Ethics* and the *Politics* helps to neutralize the eudaimonism objection but opens the virtue ethicist up to a new objection. In section 5.6 I consider the objection that this way of understanding Aristotle's

---

<sup>201</sup> From this point on, besire will be used as a generic term for any state with both a descriptive and a directive side.

project leads to an unacceptable circularity. The reason for this concern is that the preferred political structure gets its justification from the fact that it promotes the creation and support of virtuous individuals rather than from an outside source. I employ an analogy with mechanical systems to show that there are outside grounds to provide a justification for a virtue-based approach coupled with a political system designed to support the virtues. Section 5.7 provides a brief summary of the chapter as a whole.

## 5.2 The Moral Problem revisited

As a reminder, here is the moral problem as Smith formulates it.

1. Moral judgments of the form 'It is right that I  $\phi$ ' express a subject's belief about an objective matter of fact, a fact about what it is right for her to do.
2. If someone judges that it is right that she  $\phi$ s then, *ceteris paribus*, she is motivated to  $\phi$ .
3. An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences.<sup>202</sup>

Each proposition is independently plausible, Smith claims, but they have seemed collectively implausible. Central to this problem is a concern about how facts and motivations in the moral realm can exhibit such a regular connection. The first claim concerns the apparent objectivity of moral judgment. It may seem possible to satisfy (1) and adopt an error theory as J. L. Mackie does.<sup>203</sup> Mackie maintains that when ordinary people make moral judgments they are making claims about the existence of objective values but that there are no objective values. People mistakenly believe that there are moral properties even though there are none. As a result, such claims wind up being false but they do have objective truth values. However, it is clear that Smith intends something more. Smith summarizes his views on this by saying the following: "we seem to think that moral questions have correct answers (and) that the correct answers are made correct by objective moral facts."<sup>204</sup> Mackie seems to agree with Smith in

---

<sup>202</sup> Smith 1994, p. 12.

<sup>203</sup> Mackie 1977.

<sup>204</sup> Smith 1994, p. 6.

that people tend to think that moral statements are about objective facts and can be both false and true.<sup>205</sup> It is this intuition that (1) captures and that we would like to preserve if possible.

It is precisely because (1) concerns our beliefs about the existence of objective moral facts that we have trouble seeing how it and (2) can both be true. We do not generally think of beliefs as being motivating absent some independently existing desire. If (3)'s claim that beliefs and desires have independent existences is correct then it is difficult to see how both (1) and (2) could be true. If (1) is true then moral judgments are beliefs and beliefs do not typically exhibit the regular connection with motivation that we see in moral judgments. Therefore (2) should be false. But (2) seems to be true and, given that it is, we expect that (1) would be false and that moral judgments would express something other than a belief about an independently existing realm of moral facts.

The denial of the truth of (3) makes it significantly more plausible that (1) and (2) are both true. That is, the apparent conflict between (1) and (2) stems from our acceptance of (3). The solution that is on offer here entails that (3) is false, though the weaker conditional read from right to left can still be true. (If an agent has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences, then she is motivated to act in a certain way.) In chapter four of *The Moral Problem*, Smith argues against the existence of mental states that have both a content side and a motivational side. However, it was argued in chapter three of this dissertation that Smith's argument against these kinds of mental states only works if we assume that all moral mental states must be of this nature. It was argued in chapter four that Aristotle's moral philosophy can be interpreted as holding that moral motivation can be both externalist and internalist. Aristotle's account of moral education makes it plausible that there would be both kinds of motivation, especially when seen in the light of evolutionary considerations.

---

<sup>205</sup> Mackie 1977, pp. 30-35.

A successful defense of an Aristotelian-style virtue ethics can provide a naturalistic solution to the moral problem. First, claims about what it would be right to do in a situation can be derived from what the virtuous person would do when faced with that situation. For example, the statement, 'it is right to help a friend in need' is true because that is what the virtuous person would do in the situation in question. The situation is a bit more complicated than this. It would not be right to help a friend cover up a murder and the virtuous person recognizes this. It is more helpful to rephrase the statement above to make it specific to a particular act of helping, e.g. 'it is right for Jeffrey to help Mark move into his new apartment'. Second, possessing a virtue involves possessing a state that is both belief-like and desire-like. When a virtuous person recognizes that he is in a situation that calls for helping behavior he recognizes facts about various aspects of the situation. If Jeffrey is virtuous he will recognize relevant aspects of the situation (e.g. that Mark needs help moving things to his new apartment, that he (Jeffrey) is Mark's friend, etc.) and this recognition will be accompanied by a motivation to help. If Jeffrey is not so motivated then he does not possess the relevant virtue. Thus we can explain the second thesis. Third, Aristotelian virtue ethics coupled with a widely accepted psychological theory gives us a reason to reject (3). This psychological theory will be discussed in the subsequent section. Finally, it is not my intention to offer anything like a full defense of virtue ethics from all objections that might be leveled against it. That would involve (at least) a book length treatment in itself. I do defend virtue ethics from two specific objections—that all the theories of besires offered to this point entail that there are no moral facts and that possession of the virtues is not necessary for a person to flourish.

### 5.3 Objection 1 and response—three theories of besires

It is one thing to say that moral mental states can be 'besires' but is there any evidence that such states actually exist? If there is not then the naturalist appears to have a problem meeting the first two of Smith's three theses. If moral claims do not express the subject's belief

about an objective matter of fact then (1) is false and the result will be some form of anti-realism. Fortunately, there are three recent theories of mental states that posit just such states though they do not call them *besires*. In the next section, I will briefly survey these three theories and argue that evidence from psychology suggests that some such theory is probably true. However, I will remain agnostic on the question of which, if any, of the theories is correct.

Dennis W. Stampe argues for a theory of desire on which desires, like beliefs, are representational states.<sup>206</sup> Initially this seems implausible. Beliefs are generally taken to be states that represent the way the world is whereas desires do not. Generally, we desire something that is not the case. If I want to earn my Ph. D it is not likely that I already have my Ph. D. If I did then my desire would be satisfied. However, Stampe's theory of desire sees desire as a state the purpose of which is to aim at the good or needs. Stampe suggests that we can explain desires that are not aimed at needs as "a misrepresentation of a need."<sup>207</sup> In Stampe (1987), Stampe argues that desiring is a kind of perception.<sup>208</sup> Oftentimes what we perceive is a bodily state in which what is represented is represented as good (e.g. it would be good if I were to have a drink of water). An attractive aspect of Stampe's view is that his approach is Aristotelian so it fits well with the project of this dissertation. Further, Stampe's account of desire builds in a representational aspect as well. We can think of what I will call a Stampe-desire as having a content (e.g. 'it would be good for me to do A') and also a mode of representation (e.g. 'I will do A'). As a result, Stampe's desires wind up being a kind of *besire*.

Another kind of mental state having the structure of a *besire* is what Tamar Szabó Gendler dubs *aliefs*.<sup>209</sup> Like beliefs, *aliefs* include a factual or representational side but are also affective and behavioral. Gendler describes paradigmatic *aliefs* in the following way.

A paradigmatic *alief* is a mental state with associatively linked content that is representational, affective, and behavioral, and that is activated—consciously or

---

<sup>206</sup> See Stampe 1986 and 1987.

<sup>207</sup> Stampe 1986, p. 168.

<sup>208</sup> Stampe 1987, p. 359.

<sup>209</sup> See Gendler 2008a , 2008b and 2012.

nonconsciously—by features of the subject’s internal or ambient environment. Aliefs may be either occurrent or dispositional.<sup>210</sup>

Some aliefs are belief-accordant, meaning that there is no conflict between the behaviors generated by the alief and what we expect from a belief in the circumstances. In such circumstances, explanations of behaviors can often be generated by appeal to standard belief/desire explanation models, though this does not mean that such explanations accurately capture what is going on in every belief-accordant situation. This becomes apparent when we consider a situation in which we appear to alieve something different than what we believe. Here is one example that Gendler provides. An adult movie-goer watching a horror movie screams and clutches at his seat when it looks as if green slime projected on the screen is headed straight toward him. It would be unusual to say that the movie-goer actually believes that he is about to be attacked by green slime. He knows that he is in the movie theater and that he is not being threatened by green slime. If asked, he will say as much. But his behavior indicates the presence of an alief with roughly the content: “Dangerous two-eyed creature heading towards me! H-e-l-p . . . ! Activate fight or flight adrenaline now!”<sup>211</sup> The fact that alief and belief can come apart gives us reason to believe that there may be two largely independent systems at work.

The tripartite structure of an alief gives us reason to think that some moral mental states may fit this model as well. Gendler tells us that aliefs can be either innate or habitual and stresses the importance of bringing our aliefs in line with our beliefs.<sup>212</sup> She recommends Aristotle’s method of bringing our habits (aliefs) in line with our actual moral beliefs as one of two ways of accomplishing this.<sup>213</sup>

---

<sup>210</sup> Gendler 2008a, p. 642.

<sup>211</sup> Gendler 2008a, p. 637.

<sup>212</sup> Gendler 2008b, p. 552 and section 4.

<sup>213</sup> Gendler 2008b, section 4.1 and 4.2.



Ruth Millikan provides a detailed theory of representation that prominently includes mental states that have a structure which is similar to that which Gendler provides.<sup>214</sup> Millikan calls these states *pushmi-pullyus* and embeds them in a theory of signs, both intentional and otherwise. A natural sign which is not intentional does not have the purpose of transmitting information though it may do so if there is a “consumer” who can read the signs. (A consumer can be thought of as a mechanism the purpose of which is to read signs.) That is, we learn to read natural signs which do not have any purpose. For example, black clouds in the sky may be a sign of rain and we can learn to read these signs. However, the purpose of an *intentional* natural sign is to be read. Thus we can think of such a sign as having a producer who created it and a consumer whose job it is to read the sign.

Millikan identifies three different kinds of mental states.<sup>215</sup> The first is a descriptive intentional representation. The purpose of this type of representation is to be isomorphic with the affair that it is supposed to represent. This kind of representation corresponds to what we would usually call a belief. When such a representation occurs in language, it generally has the form of an indicative sentence (e.g. the chair is brown). The second is a directive intentional representation. With this type of representation, it is the consumer’s job to alter the world to match the representation which the producer has created. In language, this kind of representation frequently takes the form of an imperative sentence (e.g. clean your room). The third type of representation is what Millikan calls a *pushmi-pullyu*. This type of representation is somewhat like a descriptive and directive representation in one mental state in that it both describes the way the world is and directs the consumer to act in some way to alter the world.

At this point it should be clear that, like an alief, a *pushmi-pullyu* might be the kind of mental state that is structurally isomorphic to a value judgment that has both a descriptive and a directive aspect. However, there is an important kind of disanalogy between these kinds of

---

<sup>214</sup> Millikan 1995 and 2004 as well as other places.

<sup>215</sup> Millikan 2004, chapter 6.

mental states and virtue. To help bring out this disanalogy and why it could cause a problem for the project of this dissertation, it will be helpful to discuss Millikan's explanation of the nature of natural signs and how we track them. Millikan defines natural signs as "structured world affairs."<sup>216</sup> Included in the structure of a natural sign is the time and place in which it occurs but the structure may include other factors as well. For example, if the natural sign in question is a track in the mud then the size, depth, and angle of the track relative to the other tracks may also be part of the structure.<sup>217</sup>

Encountering such a sign is not necessarily informative. Our ability to read the sign depends on the recurrence of "the same sign-signified relation."<sup>218</sup> For example, if wherever we see a certain type of track in a certain woods that track is a sign of pheasant we can learn to associate those tracks in that location with the presence of pheasant. If similar tracks are made by a different kind of bird, but only in the spring and summer and only by a pheasant in the fall and winter then these tracks will be a sign of pheasant when they occur in the fall and winter but not in the spring and summer.

Stampe, Millikan and Gendler's accounts seem structurally adequate when it comes to providing an explanation of the structure of desires. Stampe's account is Aristotelian and fits well with the Aristotelian account offered in this chapter. Millikan offers her account of pushmi-pullyus as part of a larger teleological theory of mental representation. It may be argued that if we reject her overall theory then we are free to reject her account of pushmi-pullyus as well.<sup>219</sup> However, Gendler's theory may be preferable. Gendler's account appears to be less vulnerable as it is offered as an attempt to explain belief-discordant behavior in which we may claim to believe one thing but, in some ways at least, behave as if we believe something else. Her

---

<sup>216</sup> Millikan 2004, p. 47.

<sup>217</sup> Millikan 2004, p. 48.

<sup>218</sup> Millikan 2004, p.49.

<sup>219</sup> I do not mean to suggest that we *should* reject Millikan's overall theory.

theory appears to be independent of any particular ideological approach. I will not be concerned with debating the merits of these theories and will not choose one over the others.

Fortunately, there appears to be a significant amount of support in the field of psychology for the presences of besires, regardless of which theory we settle on. Many in psychology appear to accept what is sometimes called the two-system theory of cognitive processing.<sup>220,221</sup> Though I cannot go into depth about the theory here I wish to sketch an introduction to it and show how it provides prima facie support for besire-like theories. (I intend to explore the relevance of the two-system theory to the presence of besires as well as its relevance to virtue ethics in future work.) Daniel Kahneman describes the two systems in the following way:

System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control. System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of system 2 are often associated with the subjective experience of agency, choice, and concentration.<sup>222</sup>

System 1 operates in such a way that we are frequently unaware of its operations and the effects it has on our behavior. For example, we can become primed in various ways by seeing certain images or reading words or text and this can affect our behaviors in ways in which we are completely unaware. What is especially relevant to our discussion is that System 1 seems to involve what is sometimes called ‘hot cognition’ whereas System 2 involves ‘cool consideration’. This is also sometimes referred to as the automaticity of affect. When System 1 is at work, it seems that there is an automatic connection between perceiving that something is the case and being motivated to respond in some way. For example, it has been argued that automaticity of affect occurs when we think about politicians, political issues and political

---

<sup>220</sup> Not everyone agrees with the two-system theory. See Keren and Schul 2009 for a critical examination of this approach.

<sup>221</sup> Two-system theory is also sometimes called dual-processing theory. An introduction to two-system theory can be found in Kahneman 2011.

<sup>222</sup> Kahneman 2011, pp. 20-21.

parties.<sup>223</sup> On the other hand, system 2 thinking is not considered to be affect-laden. One intriguing possibility is that the connection between moral beliefs and moral motivation may depend on how the information about morality or any morally-relevant situation is processed. If it is processed by System 1 then we should expect that there will be motivation accompanying the moral beliefs but if it is processed by System 2 then there won't.<sup>224</sup> It may be that the process of inculcating the virtues that Aristotle describes involves creating an associative link between features of the environment and a response pattern. This will also be explored in future work.

#### 5.4 Objection 2 and Response—A Virtue Ethics Approach

It seems that we possess a number of theories positing states with a desire-like structure. Further, the existence of such states appears to receive a significant amount of support from work done in psychology. This concludes my work to establish the plausibility of the first of the two claims that I set out to prove in this chapter—that there are mental states that have the structure necessary for there to be desires. However, it could be argued that the plausibility of (especially) Millikan and Gendler's theories supports the first claim but actually undermines any attempt to establish the second—that there are objective moral truths.<sup>225</sup> Here is the concern. In Millikan's theory, we track signs of something that has been, or is, present in a particular domain. The tracks in the woods are a sign that pheasant are there or have been there recently (as is the sound that pheasant makes or the pattern of light impinging on the retina when we see a pheasant). Similarly, when we have an alief that we are about to be attacked by green slime though there is no actual green slime, there is green slime that we believe is present. That is, we have an actual visual representation of green slime. Can such an

---

<sup>223</sup> See Lodge and Taber 2005.

<sup>224</sup> I intend to explore the relevance of the two systems to moral motivation in future work. Wong 2009 provides an interesting discussion of the role of system 1 processes in moral motivation.

<sup>225</sup> I exclude Stampe's theory here because it is Aristotelian.

account be extended into the moral realm to claim that we track moral properties, such as goodness or rightness, in a way similar to the way that we track pheasant or would track (real) green slime? Sharon Street raises serious concerns regarding such an account. Street poses the following dilemma for moral realists and other value realists.

On the one hand, the realist may claim that there is *no* relation between evolutionary influences on our evaluative attitudes and independent evaluative truths. But this claim leads to the implausible result that most of our evaluative judgments are off track due to the distorting pressure of Darwinian forces. The realist's other option is to claim that there *is* a relation between evolutionary influences and independent evaluative truths, namely that natural selection favored ancestors who were able to grasp those truths. But this account, I argue, is unacceptable on scientific grounds. Either way, then, realist theories of value prove unable to accommodate the fact that Darwinian forces have deeply influenced the content of human values.<sup>226</sup>

Street believes that it is implausible to think that evolution had no effect on our evaluative attitudes. Many of the evaluative attitudes that we have are ones that we would expect evolution to select for, as beings with these attitudes would be ones that were more likely to get their genes passed down to subsequent generations than would beings with contrary attitudes. Many of our evaluative attitudes promote our own survival and the survival of our kin either directly or by promoting beneficial relations with non-kin, and we are much more likely to have these evaluative attitudes than we are attitudes that undermine the survival of ourselves and our kin. However, if the realist denies a relation between evolutionary forces and evaluative truths then she or he must maintain that any coincidence between evolutionary forces and evaluative truths is mere good fortune and is not to be expected in most cases. Street employs the following useful analogy. Suppose that we were to set sail for Bermuda but rather than charting a course and making an effort to maintain that course, we let ourselves be guided by the wind and the tides. If we just so happen to wind up in Bermuda, this will be coincidental. We should not expect to wind up in Bermuda and we should be surprised if we actually do wind up there. Likewise, if there is no relation between evaluative truth and evolutionary forces, the moral

---

<sup>226</sup> Street 2006, p. 109. All subsequent references to Street will be to this paper unless otherwise noted.

theorist who insists upon realism should not expect that evolutionary forces will lead us to evaluative truth.

However, the close connection between the evaluative attitudes that we would affirm as true and the actual evaluative attitudes that evolution has selected for becomes surprising on a realist account if there is no connection.<sup>227</sup> Street gives a list of evaluative judgments that we tend to make followed by a list of possible judgments which are widely rejected.<sup>228</sup> The actual evaluative judgments that we make are ones that we should expect people to make if the theories of natural selection, kin selection and reciprocal altruism are roughly true. On the other hand, the judgments that we do not tend to make are ones that we would expect evolution to weed out when they occurred among our ancestors. The best option for the realist then seems to be to embrace the other horn of the dilemma and maintain that there is a relation between evolutionary forces and the contents of our moral beliefs.

The next step is to explain the relation. The obvious option is that the relation is a tracking relation. The realist can maintain that we track moral truths and that our beliefs are an outgrowth of this tracking process. Sometimes things will go amiss but generally our moral attitudes will be true.<sup>229</sup> The realist need not maintain that we have evolved a specific ability to track moral facts. Rather it may be that we evolved a different ability and that our ability to track moral facts piggybacks on this other ability. In a response to Street 2006, David Copp points out that evolution did not select for an ability to have beliefs about the origins of the universe

---

<sup>227</sup> Street favors an anti-realist account.

<sup>228</sup> See pp. 115-116 for Street's lists.

<sup>229</sup> Strictly speaking, it might be better to say here that our tracking of moral properties will be successful often enough to explain why our representations of moral truths really are such representations. In certain theories of representation, such as Millikan's, our representations need to get it right just often enough for them to perform their function. For example, suppose that we make moral judgments about the presence of wrongness but that most of these judgments are mistaken because there is no wrongness. If the small percentage of correct wrongness judgments explain why we make wrongness judgments, perhaps because identifying wrongness when it is present is very important, then this will be sufficient to explain why the purpose of our wrongness judgments is to represent wrongness. Representations can be wrong much of the time as long as they are right often enough. Similarly, very few sperm actually manage to fertilize an egg but the purpose of sperm is still fertilization.

even though we are capable of having such beliefs.<sup>230</sup> However, we still need an explanation of why we should believe that the ability to learn moral facts is an outgrowth of some other ability which was selected for. A specific ability to recognize tables and chairs was not selected for but an ability to recognize medium-sized objects was plausibly selected for because doing so provided a survival advantage to our ancestors.<sup>231</sup> The inability of moral realists to provide a similar explanation concerning our moral beliefs in no way proves that realism is wrong but it does cast doubt upon this sort of explanation. Unfortunately for the moral realist, there appears to be a stronger account.

Street introduces the *adaptive-link account*. The adaptive-link account claims that evolution's having selected for evaluative attitudes which will have a tendency to lead to evaluative judgments can be thought of as mechanisms that "serve to link certain kinds of circumstances . . . with adaptive responses."<sup>232</sup> Street is clearly arguing for a functional account of morality. She compares evaluative judgments to reflex responses and says that though "there are radical differences between the (two) . . . there is a deep analogy between their functional roles" (128). Street raises three interrelated points to show that her anti-realist functionalism has a number of advantages over the tracking account. First, it is more parsimonious because it can explain our judgments without the need to postulate independently existing entities. A second and related point is that the tracking account suffers from a lack of clarity that is not problematic for the adaptive-link account. The main point here is that it is unclear why evolving an ability to track independent evaluative truths would be advantageous. Finally, unlike the tracking account the adaptive link account does a good job of actually explaining why we would make the evaluative judgments that we do—making the kinds of judgments that we do provides a survival advantage.<sup>233</sup>

---

<sup>230</sup> See Copp 2008, p. 190.

<sup>231</sup> See Street 2008, pp. 216-217 and 2006, pp.142-144.

<sup>232</sup> Street 2006, p.126. Street's claims here are consistent with an account of virtue ethics.

<sup>233</sup> Street 2006, pp. 129-132.

If Street is correct then we can explain why we have the “moral” responses we do to certain situations not by citing a moral property to which we are reacting but simply by citing a link between situations and responses. In some cases, the link may be innate and in other cases we may acquire the link because we live in a society that encourages a certain behavioral response in a given situation. Yet, it is not clear that we must embrace Street’s skeptical conclusion. We could try to respond by adopting a hedonistic theory of value and claim that our moral judgments track pleasure. We can reliably say which kinds of behaviors will tend to be pleasure producing and which will have the opposite effect. Notice also that were we to pursue such an approach, there would be no need to do so in a simple-minded way. We are free to think carefully about the consequences of our behaviors, becoming aware of situations in which an immediate pleasure might be followed by pain or vice versa. There are well known problems with these sorts of approaches to ethics and most consequentialists do not focus on pleasure exclusively.

It also seems plausible that we could track properties that are conducive to group flourishing. Again note that we are not obligated to adopt a simple-minded view of flourishing in which we do not take account of long term consequences. Neither Gendler nor Millikan gives us a normative account of which behavioral responses ought to be linked to which situations. However, we can draw upon virtue ethics for this. The virtue ethicist tells us that people ought to respond in such situations in the way that a virtuous person would respond. If the person is virtuous, then her response will be one that leads to both group and individual flourishing.

One way of understanding Aristotelian virtue ethics involves familiarizing ourselves with domains. Just as I can come to recognize that certain signs within a certain domain show that pheasant are present, certain signs within a domain may show me that a virtue such as generosity or temperance is called for.<sup>234</sup> Rather than draw the inference ‘X is present here’, I draw the inference ‘Y is called for here’, where Y represents a response of some sort. Further,

---

<sup>234</sup> See McDowell 1979, section 2, 331-333.



the sort of response which is called for in a particular situation is not the only possible response. For example, suppose that I have a neighbor who is struggling financially. I can look at my neighbor's situation and see it as providing the opportunity to exploit as well as the opportunity to be generous. I may offer my neighbor much less for his car than I know it to be worth hoping that he will accept the offer out of desperation. Yet the virtue ethicist can counter that there is an objective fact about which opportunity I ought to perceive. If I am virtuous then I will see the situation as calling for generous rather than exploitive behavior.

Further, the virtue ethicist can argue that the character traits that we call virtues are justified because they are a necessary component for their possessors to live the good life. If I wish to live a eudaimon life, and Aristotle tells us this is a goal that we all share, then it is in my best interest to acquire the virtues. Having done so, I will see my neighbor's situation as one that calls for generosity and I will not see it as an opportunity to exploit my neighbor. It would not be rational for me to exploit my neighbor as this would frustrate my end of living the eudaimon life.

### 5.5 Objection 3 and response—An Aristotelian Justification of Moral Objectivity

Aristotle's view of the ethically ideal person was surely affected by his overall teleological views of biology. As a fully functioning hand has the ability to grasp things and a fully functioning eye has the ability to see things, a fully functioning human being will possess certain behavioral characteristics (i.e. the virtues). Aristotle recognized that the possession of these characteristics was not essential for membership in the species 'human being'. Many people suffer from moral continence or incontinence. However, the person lacking the virtues would be like an eye that could only produce blurry images or a hand in which the fingers had limited mobility resulting in things frequently being dropped. The person is still a person but is malfunctioning.

It seems appropriate here to say something about the relation between Aristotle's essentialism and issues of sex. Aristotle has (rightly) been taken to task for his views on women.<sup>235</sup> I will mention just one of Aristotle's views. In *Politics*, Aristotle tells us that women have the deliberative faculty (i.e. reason) "but it is without authority" (1260a 14). If a woman's reason is lacking in authority then she cannot be virtuous. Fortunately, we need not accept such claims from Aristotle. Assuming we can defend an Aristotelian-style ethics sans the blatant sexism there is a related issue here concerning any form of biological essentialism as applied to human beings, regardless of whether there are essential differences between men and women.<sup>236</sup> Is it reasonable to assume that there is one form of ideal functioning that fits all human beings so that there will be one right list of virtues that will fit all human beings? Really the question here concerns whether there is any common human nature. There is clearly variation between the ways persons in different cultures live their lives.

Martha C. Nussbaum defends a form of essentialism in which she identifies a number of human qualities and capabilities that are common to human life.<sup>237</sup> I will not try to defend the details of Nussbaum's view here but would like to draw attention to two important aspects of it. First, the qualities and capabilities that she mentions are ones that are, by and large, open to empirical confirmation or refutation. For example, she points out that we all have need for food and drink and shelter. If we were to find human beings who did not share these needs then we would need to rethink whether these qualities were essential parts of, as Nussbaum puts it, a *human form of life*. Second, her view of human functions, which she calls *the thick vague theory of the good*<sup>238</sup>, leaves open the possibility that there may be multiple social structures in which human beings can function as they should.

---

<sup>235</sup> See e.g. Matthews 1986.

<sup>236</sup> For a recent discussion of gender essentialism, see Witt 2011.

<sup>237</sup> Nussbaum 1992.

<sup>238</sup> Nussbaum 1992, p. 214.

Aristotelian virtues are those character traits that are necessary for human flourishing. As such, Nussbaum's theory appears to leave open the possibility that there may be more than precisely one set of virtues that best leads to overall flourishing given the possibility of multiple social structures which will allow human beings to function as they should. This leads to the possibility that there may not be one agreed upon set of character traits that are identified as virtues as different cultures may allow people to manifest their functioning in different ways. Here are two examples of this. First, we all need to eat food but the types of food available vary from place to place. As a result, we will nourish ourselves with different foods and will develop different ways of obtaining and preparing that food. For some that will mean engaging in hunting and gathering behaviors, for others it will involve a trip to the grocery store, and still others will go to a restaurant and pay others to take care of this for them. Second, we all have a need for some kind of affiliation with other human beings though the forms this takes can vary. We may share a living space with another human being or we may interact with that human being from a great distance over electronic media. Though it is not directly relevant to the main project of this dissertation (i.e. providing a solution to the moral problem), presumably we have some ability to measure how well different practices allow us to function and meet our needs. Nutritional deficiencies have health consequences and can be measured. Assessing the success or failure of different ways of meeting our need for affiliation may be less straightforward but not impossible. People who are failing to meet these needs may show higher degrees of depression and be less healthy than others. These tasks would seem to fall largely into the field of psychology and biology.

Even if we assume that there will be multiple sets of practices that promote human flourishing, it does seem as if there would be some overlap in the virtues. Temperance, for example, is a necessary virtue in a wide range of environments. There are many ways in which a person can overindulge in pleasure or indulge in pleasure in the wrong way. Contra Aristotle,

Sarah Conly argues that the virtues are not a necessary condition for living a flourishing life.<sup>239</sup> Conly gives examples of flourishing without the virtues of courage and justice. She argues that we can imagine a species called the puppeteers, who though much like human beings, will not ever risk their personal security for others. (The puppeteers are a fictional species that occur in some of the works of Larry Niven including *Neutron Star* and *Ringworld*.) Essentially, puppeteers are cowards but they are able to have successful relationships with other members of their species, as there is no expectation that one ought to sacrifice oneself for others. They also possess the technology to protect themselves from other species that might be hostile and have developed a policy of retaliating against those who harm a puppeteer by leveling devastating economic sanctions against the people in question. It seems as if members of such a species could flourish even though they lack courage.

I do not find this counterexample particularly convincing. The virtues are intended to be a necessary condition for *human* flourishing and the puppeteers are not human. Could a human society like this exist for an extended period of time? This is a difficult question to answer because it presupposes a solid grasp of human nature and the extent to which this nature is malleable. However, I do not think that we need to be able to answer this question. The fact is that courage appears to be a character trait that benefits its possessor in the social situations that we human beings face today. It is all too common to encounter a person who is aggressive and mean. Though there are times when it is appropriate to back down in such a situation, there are other times in which there is a good that might reasonably be achieved by standing up to a bully.<sup>240</sup> There are other situations in which the threat is not from another person, at least not directly, but comes from a situation that may arouse fear, such as public speaking, in which one fears the judgment of others, or a midlife career change in which one gives up what one knows to pursue a different career.

---

<sup>239</sup> Conly 1988.

<sup>240</sup> Sometimes just putting the bully in his or her place may be the good which is achieved.

Conly's second example concerns the virtue of justice. Conly states that justice "involves respect for others and a proper estimation of their worth, and acting in a way which reflects this."<sup>241</sup> She goes on, "Can we flourish without justice? If circumstances are suitable, it seems we may."<sup>242</sup> A farmer, she tells us, may flourish in a just culture by taking advantage of the misfortunes of another by acquiring his equipment at a discount. Further an unjust person, like Lorenzo the Magnificent (Lorenzo de Medici), may flourish in a culture in which there is no expectation that people will behave justly. What is necessary, she tells us in footnote 18, are a set of coordinating rules that people can use to guide their behavior and that justice is not the only source for such rules. Conly draws the conclusion that the moral justification for preferring the character traits that we call virtues will need to be justified in terms other than their necessity in the life of a flourishing individual.

Conly's argument regarding the non-necessity of justice is stronger because it cites examples of real people and realistic situations. The advocate of eudaimonism can dig in her heels and insist that the farmer who exploits his neighbor and Lorenzo de Medici are not flourishing, regardless of their outward appearances. Comforting though this thought may be, it seems to run counter to our experiences and seems hard to justify empirically. Even a strong correlation between a person being virtuous and that person flourishing is not sufficient to show that the virtues are necessary for flourishing as the correlation of the two may be due to living in an environment that rewards virtuous behavior.<sup>243</sup> A related concern is that if de Medici insists on behaving justly even though he finds himself in an unjust culture he may be judged to be behaving irrationally. To behave virtuously he must sacrifice his own flourishing. The necessity of the virtues in a flourishing life is supposed to make the possession of, and consequent acting upon, the virtues rational because it contributes to the individual's flourishing.

---

<sup>241</sup> Conly 1988, p. 91.

<sup>242</sup> Conly 1988, p. 92.

<sup>243</sup> Strictly speaking, a perfect correlation would not prove that the virtues were necessary for flourishing either but the conclusion that they are would presumably be justified on inductive grounds.

The eudaimonist may try to appeal to the role that the noble (fine) play in Aristotle's ethics. Virtuous persons get pleasure from behaving virtuously because virtuous behaviors are noble. The non-virtuous person does not have access to this realm of pleasure. If noble pleasures are a necessary condition for living a flourishing life then Lorenzo de Medici and other non-virtuous people are not flourishing, regardless of how things look. The problem with such a response is, as was mentioned above, it is hard to justify to one who is skeptical. It is not uncommon to derive a positive feeling from having behaved rightly but this is a long way from cashing out just what such noble pleasures are like and why they are necessary for flourishing. In the following section I will draw upon evidence from Aristotle's *Nichomachean Ethics* and *Politics* to argue that Aristotle's philosophy does provide the resources to develop a response to the eudaimonism objection.

I think that it is quite difficult to give a justification of Aristotle's virtues in terms of their contribution to the good life of the individual that possesses them independent of the circumstances in which the agent finds himself or herself. It may be a mistake to try as it is not clear that Aristotle ever intended this. There are two points about Aristotle's moral philosophy that seem of particular relevance, the deep connection that he sees between moral philosophy and political philosophy and the conception of human beings as creatures for whom it is their nature to form larger groups. I begin with the first.

The connection between what we would call moral philosophy and political philosophy is apparent in both the *Nichomachean Ethics* (NE) and the *Politics*. In the second chapter of the first book of NE he discusses the connection between the good for an individual and the highest good, which is the good for a city-state. He assigns all "sciences concerned with action"(1094b 5) a subservient role to political science and concludes that they are a form of political science. Aristotle ends the second chapter of the first book of NE with the following words. "And so, since our line of inquiry seeks these [goods, for an individual and a community], it is a sort of political science" (1094b 11-13).

Another indicator of the importance Aristotle places on the connection between ethics and political philosophy comes from the way NE ends. In the final chapter of the final book, Aristotle turns his attention to the sorts of social and political structures that will be necessary for developing virtuous individuals. In addition to emphasizing correct teaching, he emphasizes the role of law in the development of virtuous individuals. He states that “it is difficult, however, for someone to be trained correctly for virtue from his youth if he has not been brought up under correct laws” (1179b 32-34). Nor are laws necessary for just the young. Aristotle continues, “they (the young) must continue the same practices and be habituated to them when they become men. Hence we need laws concerned with these things also, and in general with all of life” (1180a 2-5). From these two passages, it seems clear that Aristotle sees the development of correctly functioning human beings as taking place in a society with the right kinds of structures in place, including the right laws.

It is further helpful that, in the best kind of state, the ends of the individual and the ends of the state are not in conflict. That is because the best kind of state will be that which conduces to the best kind of life for the individual. Aristotle says in the *Politics* that, “We ought therefore to ascertain, first of all, which is the most generally eligible life, and then whether the same life is or is not best for the state and for individuals” (1323a 19-21). Following a discussion of the necessity for virtue, some external goods and good health for the good life, Aristotle concludes that “the best life, both for individuals and states, is the life of virtue, when virtue has external goods enough for the performance of good actions” (1323b 40-1324a 2). As is the case with individual virtue, some amount of external goods will be necessary for the state to be virtuous. A state which aims at ruling over the greatest amount of people or acquiring the greatest amount of wealth will not be the ideal state for virtuous individuals.

A second important point about Aristotle’s moral philosophy is that he understands it to be in the nature of human beings to be highly social. In the second chapter of the first book of the *Politics* he describes how the city-state develops from smaller structures. Human beings

naturally form families and these families naturally form into villages. When several villages unite, the city-state is created and the city-state is natural because it is an outgrowth of these earlier structures. Though he is very vague concerning the process of moving from family to village to city-state, he does imply that we cannot be self-sufficient, or at least that it is harder to be self-sufficient, in the smaller structures.<sup>244</sup> The original purpose of the state is to provide for our basic needs and then it exists to provide an opportunity for human beings to live a good life.<sup>245</sup> He concludes that “it is evident that the state is a creation of nature, and that man is by nature a political animal” (1253a 1-3).<sup>246</sup>

NE focuses in large part on the development of the individual but it does pay quite a bit of attention to providing practical advice on how we are to interact with others. The NE contains two books on friendship and it is clear that Aristotle does not envision people living in isolation from others. The word that he uses is ‘philia’ and this term has significantly wider scope than the term friendship would suggest. It appears to describe just about any relationship of good will between two people including relationships between family members and relationships for the purpose of mutual benefit, as in a business relationship. The NE also has a book devoted to justice. Aristotle recognizes the importance of each of these subjects even though, and he recognizes this too, neither fits the model of what it is to be a virtue very well.

By treating moral philosophy and political philosophy as two perspectives on the same issue as Aristotle did, we may be able to discover a response to the eudaimonism objection. The eudaimonism objection claims that it is false that the possession of the virtues is a necessary condition for the flourishing of the individual because a vicious person, such as Lorenzo de Medici, may flourish even though he was unjust and the society within which he came to rule was unjust as well. Let us reformulate the eudaimonism claim.

---

<sup>244</sup> 1252b27-30.

<sup>245</sup> Ibid.

<sup>246</sup> He reemphasizes this point at 1278b 15-25.



Eudaimonism<sub>R</sub>: An individual within the best kind of state flourishes  $\equiv$  an individual possesses the virtues, moderately good health, and a reasonable amount of wealth.

This revised definition of eudaimonism relativizes the necessity of virtues in a flourishing life to living within the best kind of state. Concerning the best kind of state Aristotle says in the *Politics* that “it is evident that the form of government is best in which every man, whoever he is, can act best and live happily” (1324a 24-25). And this means that the kind of governing that will take place is one based on just laws and practices. Aristotle makes this clear earlier in the chapter from which the quote above originates by pointing out that those who think the good life for an individual consists in the acquisition of wealth will think that the state that makes pursuing wealth its number one priority is best whereas “those who value most highly the life of a tyrant deem that city the happiest which rules over the greatest number” (1324a 10-11). Since being virtuous, possessing moderate wealth and having good health is essential to living the eudaimon life for an individual, the government should be set up to promote these goods. In such a state, Lorenzo de Medici would not flourish. To do so, he would need to undermine the laws and institutions of the state, essentially destroying the virtuous state in order to create an environment in which to flourish.

Before concluding this section it will be useful to note one other benefit that comes about from treating the moral agent as situated within a supporting political structure rather than as a moral entity to be understood in completely independent terms. The revised version of Aristotelian ethics we have been considering does much to resolve the tension between doing what is right and doing what is rational. First, it is frequently in the individual’s best interest to do what is morally right. People who do immoral things tend to alienate others and suffer from this alienation. Second, Aristotle tells us that the virtuous person takes pleasure in doing fine actions so they get something out of it, regardless of how others might be encouraged to treat them. In a just (and effective) political structure, a de Medici will be best served by behaving

virtuously, or least as virtue would dictate. Yet cases like de Medici's seem to show that there will be times when doing the right thing will not be rational.

One might object that this does not go far enough to resolve the conflict between doing what is moral and behaving rationally. It may be useful to consider the argument that Peter Railton makes for sophisticated consequentialism.<sup>247</sup> A sophisticated consequentialist is committed to some form of consequentialism as the right account of the good but is not committed to making his moral decisions in any particular way. Specifically, he need not commit himself to trying to maximize the good with each of his decisions. As Railton puts it, "a sophisticated act-consequentialist should realize that certain goods are reliably attainable—or attainable at all—only if people have well-developed characters."<sup>248</sup> We can rephrase the question as follows: "Should I choose a virtuous character to guide my actions or should I examine each possible choice in terms of how it will benefit me? In an unjust regime, choosing the former option is less likely to be rational as duplicitous behavior may be rewarded. However, a virtuous character may go a long way in a just political system, making the choice of developing a virtuous character the rational one.

#### 5.6 Objection 4 and Response—The Mechanical Systems Analogy

The solution just given is open to the following objection. The possession of a certain set of character traits,  $C_1$ , is justified because they are needed for an individual to flourish in a political structure,  $P_1$ , which is set up in such a way so that persons possessing  $C_1$  can flourish. The result appears to be circular justification. This opens up the possibility that there may a different set of character traits,  $C_2$ , which would justify a different political structure,  $P_2$ , because  $P_2$  allows persons possessing  $C_2$  to flourish, and so on. The following situation appears to result.

---

<sup>247</sup> Railton 1984, p. 153.

<sup>248</sup> Ibid, p. 158.

- Character set  $C_1$  is justified  $\leftrightarrow$  political structure  $P_1$  is justified.
- Character set  $C_2$  is justified  $\leftrightarrow$  political structure  $P_2$  is justified.
- Character set  $C_3$  is justified  $\leftrightarrow$  political structure  $P_3$  is justified.



Consider again Lorenzo de Medici. Suppose de Medici possesses  $C_2$  and that, possessing  $C_2$ , he will flourish in  $P_2$ . Now  $P_2$  appears to be justified because it allows de Medici to flourish.

How are we to determine which pairing of character traits and political structure is justified when the justification is internal to the pairing itself? Citing the fact that the  $C_1$ - $P_1$  pairing best promotes overall flourishing or maximizes flourishing will not be sufficient because it assumes that human flourishing is the good and that this good ought to be maximized. What we seem to need is an independent reason to accept this as the right pairing. Notice that appealing to the possibility of a plurality of pairings that maximize overall flourishing will not get us out of this problem because there will be many other pairings that do not maximize overall flourishing. Presumably, the pairing that allows de Medici to flourish given the character traits that he possessed is not one of them. What we need to do is provide an independent reason for believing that such pairings actually do promote the moral good. In what follows I try to provide such a reason.

### The Mechanical Systems Analogy<sup>249</sup>

In this section I draw an analogy between the virtue ethics of Aristotle and the functioning of mechanical systems. In this section I have two goals. First, I want to show that mechanical systems have enough in common with various aspects of our moral experience to justify a prima facie case that human beings develop moral systems that are similar to mechanical systems. Second, I will argue that embracing the existence of moral systems

---

<sup>249</sup> In this section and the sections that follow I am arguing for a kind of moral functionalism. I will assume that the primary function of morality, meaning the reason that we evolved to form moral systems, is to promote successful group living. (This does not mean entail that every social system that promotes successful group living is morally ideal.) For similar approaches, see Copp 1995, 2008 and Kitcher 2006.

provides a justification for preferring an Aristotelian-style solution to the eudaimonism objection that is similar to one I argued for in the previous section. Though my intention is to use the analogy to provide a response to the objection above, I believe that it can provide insight into many of the fundamental aspects of morality. I argue that there are four striking similarities between the behaviors of parts within a mechanical system that are considered good and the behaviors of human beings that are considered morally good.<sup>250</sup> I maintain that this gives us prima facie reason to think that morality may be understood as a system of a sort as well. After arguing for the four shared properties I will argue that there are two general rules that can be derived from the nature of systems themselves. This provides further support for the analogy. However, I am not trying to argue that moral systems are *exactly like* mechanical systems. Rather I want to show that they share enough properties to justify the analogy. The example I will focus on is the behavior of a cooling fan and how it functions within the larger system of an automobile though I believe that the general characteristics of a good cooling fan will carry over to other parts in mechanical systems. That is, the example is not special.

### The Four Properties

When the temperature of the engine coolant climbs to a certain point, a sensor in the radiator sends a message to the cooling fan.<sup>251</sup> The cooling fan starts and works to decrease the temperature of the engine coolant. The fan continues to run until the temperature of the coolant drops to a safe level, at which point the sensor sends another message to the cooling fan and the cooling fan shuts off. The fan should remain off until it receives another signal from the sensor to start again. There are four properties of behaviors of good cooling fans that are especially relevant.

---

<sup>250</sup> No such analogy exists for the virtues and this is a desired result as the virtues signify the correct functioning of a human being.

<sup>251</sup> In some cars, the sensor is located in the engine's computer.

The behaviors of a good cooling fan are ones which are *flourishing enhancing/maintaining*. By this I mean that these behaviors are ones which contribute to the continued functioning of the system. If the cooling fan fails to come on when it is signaled to do so, the engine may overheat causing significant damage to many of the system's parts and the system may cease to function at all. On the other hand, a cooling fan that runs continuously may cause the engine to run at a suboptimal temperature or the fan's engine may wear itself out so that the fan does not come on when signaled to do so.

The good cooling fan's behaviors are *purposeful*; the fan's purpose is to turn on and shut off at the appropriate times to keep the engine from getting too warm. Of course, a cooling fan gets its purposeful behaviors from its designer(s) but there are other things that a cooling fan does that are not a product of design and that are not purposeful. For example, a cooling fan will also generate some heat of its own and make a whirring noise but neither is a purposeful behavior. These non-purposeful behaviors are not judged to be good in and of themselves and can only be derivatively good as a sign that the cooling fan is performing its purposeful behaviors.

The behaviors of a good cooling fan are also *other-affecting* in that what it does affects other parts within the system. What the cooling fan does has a direct effect on at least one other part within the system—the coolant. Note that the cooling fan can contribute to the overall functioning of the larger system by playing a limited role in the system. The cooling fan's only job is to cool the engine coolant. As long as the temperature of the car's engine is below a certain point, the cooling fan need not come on. However, if the cooling fails to come on when it should the negative effect on the overall functioning of the system can be quite serious.

The goodness of the cooling fan lies in its purposefully behaving in the right way under the right circumstances. This makes the goodness of the behavior of the cooling fan *rule-compatible*. By this I mean that we could write rules for the cooling fan to follow if the cooling fan was capable of understanding and molding its behavior to such rules. (We might also write

the rule for someone designing a cooling fan—‘Design the cooling fan so that it comes on in situation X and shuts off in circumstance Y.’) The rule compatibility requires no additional argument because it is a consequence of the first three properties of behaviors for good cooling fans: being flourishing-enhancing/maintaining, being purposeful and being other-affecting as well as the limited number of situations in which a cooling fan finds itself. These three properties place limiting conditions on the types of behaviors which are classified as good.<sup>252</sup>

The four properties discussed above are clearly ones that are frequently taken to have moral relevance. However, I will not give a defense of the moral relevance of each property here. Rather I hope to have shown that there is prima facie reason to believe that there are similarities between moral systems and mechanical systems. Further similarities will be discussed below. I continue with the analogy. If we were writing rules to govern the behavior of cooling fans we could write the following two specific rules.

1. Turn on when the engine coolant reaches 185 degrees Fahrenheit.
2. Shut off when the engine coolant drops to 165 degrees Fahrenheit.

A cooling fan is much simpler than a human being and this simplicity makes it easy for us to capture ‘goodness’ in the pair of rules given above. However, we can stretch the analogy and distinguish between the behaviors of a good cooling fan and one that is not. We could say that a good cooling fan recognizes when it is in a situation in which blowing is called for and blows on the radiator to cool the engine coolant. Such a fan possesses the cooling virtue. A virtuous fan also recognizes situations in which not blowing is called for and ceases blowing if it is doing so. The relevance of the analogy does not lie in any claim that the “mental” life of a “virtuous” cooling fan is like that of a virtuous person. The relevance lies in what is arguably the overall goal. Virtue in a person makes that person such that she or he plays her or his role in contributing to the functioning of the larger social context of which she or he is a part, particularly assuming that the social context itself is a just one.

---

<sup>252</sup> Whether rules can be formulated which would completely adequately govern moral systems is an open question.

What role then will politics have in the good life for human beings? It seems that, in theory, there might not be any more that needs to be said on the subject on the good life for human beings. After all, if every adult person in a society is virtuous then, as long as the children are raised properly, the system should continue unabated. Similarly, if all the parts within a mechanical system are functioning as they ought to the system should continue to function as long as the parts maintain their physical integrity.

Aristotle rightly recognizes that more needs to be said but some of what he says may initially strike us as confusing. I return to the *Politics*. Having determined the goal of the best kind of state early in Book 7, one might expect the remainder of that book as well as the one that follows—which is the last book in the *Politics*—to be devoted to a list of moral principles along with a justification for each one, some of which might become laws in the ideal state. One would be disappointed. Instead he tells us about the ideal size and location of the state, at what age people ought to marry, and so on. At first, we may struggle to make sense of this theoretically uninteresting information which, admittedly, may have practical importance. Keeping in mind the analogy with mechanical systems, Aristotle's approach makes more sense if, having established a functioning system, Aristotle sees the goal of political science as protecting that system. Threats can come from either outside or inside the system.

Here is an example of a threat that may come from the outside. Aristotle considers whether it is best for a city-state to have access to the sea.<sup>253</sup> On the one hand, such access is beneficial for defense and to provide a sufficient supply of goods and we need a certain amount of external goods to live the eudaimon life. On the other hand, a connection with the sea means that foreigners who have not been raised in the best way may interact with the citizens, thereby introducing bad influences. Aristotle's solution is to have the sea port some distance from the state itself and have the sailors be non-citizens in order to minimize negative interactions while enjoying the benefits of being close to the sea.

---

<sup>253</sup> *Politics*, book 7, chapter 6.

Here is an example of threats that may come from the inside. In Book seven, chapter 17, Aristotle considers how children ought to be raised.<sup>254</sup> In order to rear strong and healthy children, Aristotle advises that we feed them foods that contain the most milk but little wine.<sup>255</sup> In addition, he tells us that “to accustom children to the cold from their earliest years is also an excellent practice, which greatly conduces to health” (1336a 12-14). If children are not raised properly they may not be healthy, and health is a necessary condition for living a flourishing life. The quote given in section 5.5 seems relevant as well. Recall that, as Aristotle is wrapping up the discussion in NE, he tells us that laws will be necessary not just for children but even for men raised correctly so that they may continue on what we might call the virtuous path. For those who have not been raised correctly, the laws can encourage them to behave as the virtuous person would.

The information Aristotle is providing is information that one cannot be expected to have simply by being virtuous, healthy and having moderate wealth. Though the virtuous person possesses the intellectual virtues in addition to the virtues of character, this does not mean that the virtuous person has had all the requisite experiences necessary to recognize these facts. The effects of interactions with foreigners are something we learn through experience and one can possess the virtues without having had any experience with foreigners. Neither can we learn a priori what works and does not work when it comes to raising children.<sup>256</sup>

Why does Aristotle not focus on laws or principles? I do not know the answer to this question but it may be that he thought that the most basic laws and principles would be obvious to the virtuous person. I believe that if we return to the mechanical systems analogy we may be able to discover two general rules that appear to be a consequence of thinking in terms of a functioning system. In addition to the four properties and the two specific rules for cooling fans, the following two general rules appear to result from the nature of systems:

---

<sup>254</sup> All of Book eight is devoted to this topic as well.

<sup>255</sup> 1336a 5-9.

<sup>256</sup> I put aside any evaluation of Aristotle’s opinions regarding proper child rearing.



- A. Do not harm others.
- B. Help others.

It will be obvious to anyone with the most casual interest in morality that these general rules are ones that many people would intuitively think are moral, at least when applied to people. In fact, if we were limited to only two moral rules to follow to guide our and other peoples' behaviors, these would probably be the two. To understand how these rules are generated by the nature of moral systems, it will be helpful to point out a difference between mechanical and moral systems. Unlike the parts of a mechanical system, parts of a moral system have a much richer life outside of their respective system.<sup>257</sup> Arguably, much of what a human being does has little or no influence on the moral system of which that human being is a part. In most circumstances, scratching an itch on my leg or deciding to watch a particular television program has no impact (or a negligible impact) on other people.

In contrast, the lives of the parts of a mechanical system are exhausted by the role they play within their system. But imagine that the parts of a mechanical system were more like the parts in a moral system. Imagine that when the car is parked by its owner for the night the parts all become *animated*. The parts go off to explore for the night, a la many an animated movie, only to return and get back into their respective places before the owner leaves for work in the morning. The specific rules (1 and 2) mentioned above will not be relevant for the cooling fan, nor would specific rules written for other parts being relevant, since the parts are no longer assembled. Yet, there are other ways that the parts could fail to be able to do their jobs when they return. One thing that could happen is that one part could somehow damage another part, making it unable to resume its job in the morning. So the first general rule about not doing harm will be important. But harm can come about by misfortune as well. In whatever way the harm comes about, it could result in a part being unable to do its job. Therefore, if a part in need of

---

<sup>257</sup> Here I do not mean to imply outside of any social environment whatsoever but rather outside of the behaviors which, strictly speaking, are considered moral.

help can be helped, doing so will positively contribute to the functioning of the system; hence the need for the second rule.<sup>258</sup>

Now consider moral systems. Moral systems consist of many specific rules governing which behaviors are acceptable, which are obligatory and which are forbidden. It is not clear that we would be able to generate enough specific rules to cover every possible situation that a person may find himself or herself in. The more general the rule, the greater the number of situations it can cover. It is reasonable to conclude that any sort of functioning system, including moral ones, that is composed of parts whose behaviors are not completely dictated by specific rules would need the two general rules mentioned above to maintain the functioning of the system. The main point is that uncertainty enters human social systems because the behaviors of the parts are not hard-wired in any way that makes their behavior predictable. As such, the two general rules are needed if the system is to continue on.

The four properties shared between the good behaviors of the parts of a mechanical system and good moral behaviors as well as the two general rules which seem to be generated by systems gives us *prima facie* reason, I maintain, to think that the ontological basis of morality might be profitably thought of as a literal system of a certain sort. Now we are in a position to see how the analogy with mechanical systems can allow an Aristotelian-style virtue ethics to meet the fourth objection.

Aristotle claims in the *Politics* that well-being is the main purpose of the state. He says the following:

First, let us consider what is the purpose of a state . . . men, even when they do not require one another's help, desire to live together; not but that they are also brought together by their common interests in proportion as they severally attain to any measure of well-being. This is certainly the chief end, both of individuals and states. And also for the sake of mere life (in which there is possibly some

---

<sup>258</sup> We might also need each part to follow an additional specific rule: Be back in your place within the system by morning.

noble element so long as the evils of existence do not greatly overbalance the good) mankind meet together and maintain the political community.<sup>259</sup>

Now it is evident that the form of government is best in which every man, whoever he is, can act best and live happily.<sup>260</sup>

Aristotle views the state as an entity the purpose of which is to meet the needs of its members.

People come together in a state because it can help them to meet their material needs and also provides the opportunity to live a virtuous life. The goodness of a state supervenes on the goodness of the members of the state.

If the mechanical systems analogy is successful in that there is enough similarity between the goodness of a part within such a system and the goodness of a human being then this appears to provide some insight into why Aristotle's view of the proper relation between the individual and the state is the correct one. This will become more apparent when we consider a disanalogy between mechanical systems and political systems. The ends at which a mechanical system aims typically lie outside the system itself because there is an outside designer who had certain purposes in mind in creating the system.

For example, we often attribute the predicate 'good' to systems and other artifacts which have a function (e.g. X is a good X-ray machine, X is a good rack, X is a good knife, etc.).<sup>261</sup> A natural way to understand how the adjective 'good' functions is that it tells us that what it modifies does or can be used to do what it is designed for. A good knife can be used for cutting things, a good X-ray machine can be used for taking X-rays, and a good rack can be used for tearing peoples' limbs off. It is widely agreed that there is no straightforward way of deriving moral goodness from the ability of the above mentioned artifacts to do what they were designed to do. Let's start with the first example. A correctly functioning X-ray machine can be used to diagnose various ailments. In order for the machine to function as it ought to its parts need to

---

<sup>259</sup> 1278b 15-28

<sup>260</sup> 1324a 24-25.

<sup>261</sup> See Foot 2001, especially ch. 5 and Thomson 2008, especially ch. 12 for recent attempts to explain human morality in terms of functions though each emphasizes the role of badness or defect in their analyses.

do what they were designed to do. This results in the X-ray machine functioning as it should but it also results in a good for those who are helped (because it provides information that can aid a doctor to help a patient). However, a mechanical system which is correctly functioning can also have bad results. This is because two (or more) people can have conflicting purposes. A correctly functioning rack can be used to tear a person's limbs off. In this case, the parts of the machine functioning as they were designed to leads to a bad result (someone being tortured). Finally, consider a good knife. A good knife is one that can be used for cutting but whether the knife's functioning as it ought to leads to good or bad results depends on how the knife is used. If it is used to prepare dinner then it has good results and if it is used to stab someone then the results are bad. As can be seen in the graph below, there is no straightforward connection between an artifact or system being able to perform the function it was designed to perform and the moral goodness or badness of the result.<sup>262</sup>

Artifact	Capacity to perform its given function effectively	Typical moral result
Knife	Present	Mixed
Knife	Absent	Mixed
X-ray machine	Present	Good
X-ray machine	Absent	Bad
Rack	Present	Bad

<sup>262</sup> In fact things are considerably more complicated than the immediately preceding discussion indicates. A X-ray machine that is functioning correctly could be used to repeatedly expose an individual to radiation whereas a working rack could be used to teach people about the horrors of torture. This is why I have used to the term 'typical' in the graph above. These complications do nothing to undermine the present argument.

Rack	Absent	Good
------	--------	------

This is a desired result. If we tried to claim that such an identification were possible then one could claim that an effectively-run but deeply oppressive culture was good because it functioned like a “well-oiled” rack. What we want to claim is that the goodness of the state can only be cashed out in terms of the flourishing of the individuals within that state. If we are making an attempt to measure the goodness of a state it must be reducible to the overall flourishing of the members of the state. A state which is structured in such a way that it allows one or a few members to flourish while the others languish will be an inferior state because it contains less flourishing than the virtuous state. *And this is what we can now claim because, unlike the artifacts discussed above, there are no ends outside of the functioning human system to use as a basis for a value judgment.*<sup>263</sup> The only place we can find goodness or badness must lie within the system itself. This is why it is appropriate to say that the (moral) goodness of the state must be reduced to the flourishing of the individuals within that state. Now we have an (ethical) justification for rejecting any character set/political structure pairings that do not maximize flourishing as they will fail to be as good as those that do.

The preceding discussion raises questions about the semantics of moral terms. To explore these it will be helpful to start with a discussion of a general problem which is not relevant to moral semantics only. The problem concerns to what extent linguistic expressions get their meanings from their contexts. To be clear, I have no solution to offer to this complex problem. However, examining this debate and the issues around it will provide some insight into how the mechanical systems analogy offered above can shed some light on the semantics

---

<sup>263</sup> Objections can be raised to this statement. On the view being advanced here, animals would lie outside the moral system. As a result, there is a question about whether their flourishing ought to be taken into account. We might also question whether the moral good lie in those actions that were pleasing to God. I set these questions aside for present purposes.

of thin moral terms like 'right' and 'wrong'. In addition to straightforward indexicals like 'that' and 'I', some semanticists have argued that many adjectives get their meaning from their context.

To begin, consider the following two sentences.

- (1) John is a tall. (Tj)
- (2) That chair is metallic. (Mc)

Asked to symbolize (1) and (2) we might write (Tj) and (Mc) to show that John possesses the property of tallness and the chair being referred to has the property of being metallic. A distinction is frequently made predicative and attributive adjectives. Roughly speaking, an adjective is predicative if a conjunctive analysis can be given of sentences which include it and attributive otherwise.

- (3) John is tall and John is a man.
- (4) That thing is a chair and that thing is metallic.

(3) causes problems that (4) does not. If asked why we ascribe the property of tallness to John we might reply by citing his height; he is 6'4". The skeptic might then ask if a building or giraffe of the same height would also qualify as tall. Admitting that they would not, we might conclude that the conjunction is false because John does not possess the property of tallness, simpliciter. We could clarify our statement by saying that John is tall *for a man*. When we judged that he was tall we were comparing him to the class of men, not giraffes or buildings. Upon further thought, we might decide that tall is really a relation that holds between John and the class of men though not the class of giraffes.<sup>264</sup>

- (5) John is tall *for a man*. (Tjm)
- (6) John is not tall *for a giraffe*. ~(Tjg)

We might decide that what 'tall' means is determined by the context in which it occurs and that it means something different in (5) than it does in (6). Some semanticists have come to believe that many if not all sentences require contextual input in order to be truth-evaluable. The idea is that, strictly speaking, what is written in (1) does not have a truth value; it is a proposition

---

<sup>264</sup> See Wheeler 1972 for a general discussion of attributives. Wheeler argues that many adjectives are two-place rather than one-place predicates.

radical, or partial proposition. Only when John is compared to the class of men or some other class do we have something with a truth value.<sup>265</sup> The class that John should be compared to is filled in by the context in which the sentence is uttered. I will revisit this issue shortly but now I want to discuss a similarity between the issue of contextualism in semantics and a parallel concern about moral statements.

Imagine someone growing up believing that the following sentence expresses a truth evaluable proposition which happens to be true:

(7) Enslaving children is wrong. (We)

The person assumes that anyone who acts to enslave children, regardless of where this occurs is doing something wrong. However, it is pointed out that in some African countries the practice occurs and that it has been occurring for a while. In fact, some of those who are doing the enslaving were actually slaves themselves. This suggests that perhaps (8) is preferable to (7), for reasons similar to preferring (5) to (1).

(8) Enslaving children is wrong in the USA (Weu)<sup>266</sup>

For the ethicist who wishes to reject the relativizing move from (7) to (8), the job is to identify some reason why (7) is true regardless of the context of utterance. We could try to identify some natural property of wrongness that enslaving children possesses. A problem with this strategy has been raised by Gilbert Harman.

If you round a corner and see a group of young hoodlums pour gasoline on a cat and ignite it, you do not need to *conclude* that what they are doing is wrong; you do not need to figure anything out; you can *see* that it is wrong. But is your reaction due to the actual wrongness of what you see or is it simply a reflection of your moral “sense”, a “sense” that you have acquired perhaps as a result of your moral upbringing?<sup>267</sup>

---

<sup>265</sup> See Borg 2004, ch. 1 for a good general introduction to this debate.

<sup>266</sup> I do not wish to give the impression that the contextualist interpretation of adjectives is as plausible for all adjectives as it might initially seem to be for ‘tall’ and ‘wrong’. See Thomson 2008, addendum 1 for a discussion of why the meaning of color predicates is probably not context sensitive. I will argue that ‘wrong’ and ‘right’ are probably not context sensitive either.

<sup>267</sup> See Harman 1977, p. 4.

Harman's point is that there is an important difference between the role observation plays in ethics and its role in science. He points out that the physicist posits the presence of a proton in a cloud chamber to explain the presence of the vapor trail she sees. The presence of the proton has an explanatory role to play in the observation of the vapor trail whereas it seems that moral wrongness is not needed to explain our judgment that the boy's behavior is morally wrong. We can explain our reaction to the burning of the cat by citing facts about our psychology and upbringing; we need not posit some additional property of 'wrongness'.

Nicholas Sturgeon responded to Harman's claim by citing the cruelty of the children's act.<sup>268</sup> The act of burning the cat possesses the features "of deliberate, intense, pointless cruelty."<sup>269</sup> Sturgeon appeals to character traits to provide an explanation for why we make the judgment we do in response to the burning of the cat. We need not cite an additional property of wrongness that the children's act possesses. Sturgeon makes it clear that he is starting from the assumption that there are moral facts and expresses doubts concerning whether moral facts can ever be reduced to non-moral ones.

An appeal to character traits appears to suggest a virtue ethics approach to morality. However, some have maintained that the virtues themselves stand in need of justification. Earlier in this chapter we explored the need to provide a justification for the virtues. Conly raised the concern that the virtues could not be justified in terms of flourishing because it is possible to flourish without them. She suggests that the virtues will probably need to be justified in terms of right actions.<sup>270</sup> For example, we might judge that the character trait 'generosity' qualifies as a virtue because possession of that character trait leads to generous acts and such acts are usually judged to be right.

In response to Conly's argument that it is possible to flourish without the virtues, I argued that there was an interpretation of Aristotle available which would allow us to restrict the

---

<sup>268</sup> See Sturgeon 1984.

<sup>269</sup> Ibid, p. 249.

<sup>270</sup> Conly 1988, pp. 93-94.



environment in which the person flourished to a society whose overall purpose was to promote the flourishing of its members. To bolster this interpretation I provided an analogy with mechanical systems and argued that the functioning of parts within a mechanical system has enough similarity to aspects of morality to believe that it provides us some insight into the nature of morality.

I believe that the mechanical systems analogy can provide some insight into how thin moral properties like rightness and wrongness are related to virtue. Let us return to the discussion of 'tall' and 'wrong'. In discussions of moral terms a distinction can be made between thick and thin terms. The idea is that there are character trait terms like 'courage', 'generosity', and 'cruelty' that both describe a kind of behavior and have an evaluative aspect to them; we approve of the first two and disapprove of the third. We also possess thin terms like 'right' and 'wrong'. We could also say of a person who engages in one of the first two behaviors that she did what was right and the listener will know that we have given a positive evaluation of the behavior. Yet the person may not know, specifically, what the person did. Harman's burning cat example casts doubt on whether there really is any such thing as a property of rightness or wrongness. It appears that thin moral terms are used primarily for evaluative purposes.

Treating 'tallness' and 'wrongness' as two-place rather than one-place predicates suggests that there may be no such thing as the property of being tall or being wrong. Herman Cappelen and Ernie Lepore resist this treatment of terms like 'tall' and 'wrong'.<sup>271</sup> They argue that these should be considered one-place predicates and that any content concerning a reference class for 'tall' is part of the speech act and not part of the meaning of the expression itself. They argue that it is the job of the metaphysician and not the semanticist to tell us what the property of 'tallness' is that all and only tall things possess. Likewise, similar considerations hold for other predicates such as 'being ready' or 'having had enough'.

---

<sup>271</sup> See Cappelen and Lepore 2005. They do not discuss moral predicates like 'wrong' though it seems likely that they would also treat these as one-place predicates.

However, Cappelen and Lepore actually do some metaphysical theorizing. For example, they suggest that for 'being ready' there is a property that all people who are ready share; they share the property of being ready for some project or other. As Cappelen and Lepore point out, just because the proposition 'A is ready (simpliciter)' exists does not mean that it is a proposition that anyone typically wants to express. If someone says 'A is ready' but expresses a more complicated proposition such as 'A is ready to start the race', this more complicated proposition is a result of the contribution of the context. Yet 'A is ready' just means that A is ready (for some project or other).

Judith Jarvis Thomson takes Cappelen and Lepore to task over this interpretation.<sup>272</sup> She gives an example in which 'A' states that 'B is ready' in a situation in which A knows that their audience will take this statement to mean that B is ready to go out on the job market, even though A does not believe this. However, it is true that A believes that B is ready to argue that the brand of coffee used in the coffee room ought to be changed. Thomson argues that we would tend to think that A has lied in this situation. I am not certain what my pre-theoretic intuition is in this case but I am inclined to say that A has spoken truthfully. However, I will not pursue this further. Truthful or not, it is clear that A's behavior is deceptive and almost certainly immoral as well. If 'A is ready' expresses a truth evaluable proposition in and of itself apart from context then A has spoken truly. Yet, A may also have expressed a proposition that is false when speech act content is taken into account. On the other hand, if 'B is ready' does not express a truth-evaluable proposition but only what may be called a proposition radical then A has not spoken truly. When we factor in the obvious contextual variables we will judge that A has said something false.

Returning to the discussion of 'wrongness', we could follow Cappelen and Lepore and maintain that 'wrongness' is like 'readiness' in that an action possesses the property of wrongness just in case there exists a context of evaluation in which that action is wrong.

---

<sup>272</sup> See Thomson 2010, addendum 1.

(9) Enslaving children is wrong (in some context).

Though this option is available it seems less than ideal. The reason is that not only is (9) true but arguably so are (10)-(12).

(10) Enslaving children is not wrong (in some context).

(11) A is ready (for some project or other).

(12) A is not ready (for some project or other).

We may want to allow for the possibility that there are certain actions whose truth values can vary relative to the context of evaluation (e.g. 'It is morally wrong for children not to allow their elderly parents to live with them') it seems that we do not want to allow for the possibility in other cases, such as child slavery. On the one hand, we want thin moral terms to be metaphysically robust enough to not allow for child slavery. On the other hand, we want these terms to not be so metaphysically robust that they require us to be able to observe some property of wrongness or rightness that actions possess.

The mechanical systems analogy can deliver the analysis we want. It can respect our intuition that enslaving children is morally wrong wherever it occurs and that we will not be able to observe wrongness in the way that we can observe (the effects of) a proton. The mechanical systems analogy suggests that wrongness is a relation that holds between an action and a social system relative to which the action can be evaluated. However, the social system to which it is compared does not vary. Thus we can rewrite (8) as follows.

(13) Enslaving children is wrong in the social system the purpose of which is to promote overall human flourishing ('the ideal social system' for short). (Wei)

On this analysis, thin moral terms pick out an evaluative relation but it is one in which the context of evaluation does not vary. The analysis delivers objectivity without requiring metaphysically suspect properties.<sup>273</sup>

## 5.7 Five objections to the mechanical systems analogy

---

<sup>273</sup> See Mackie 1977, part one, section nine for a discussion of problems with these sorts of properties.

### Objection 1 and response

Though I will not be able to provide a full defense of the mechanical systems analogy here, I do want to address five objections that may arise. The first objection is that there are some systems whose purpose is not to sustain themselves but rather to destroy themselves. Take for example a self-destruct mechanism. If such a mechanism is functioning correctly then it will destroy not only itself but whatever it is installed in. If the self-destruct mechanism is looked at as a system in itself then it is a system that, if functioning correctly, will destroy itself as soon as it is activated. If it is considered to be a part in a larger system then it is a part whose only function is to destroy the system of which it is a part. Doesn't this show that it is neither in the nature of systems to even continue functioning nor that a good behavior of a part is one that tends to maintain or enhance the flourishing of the system as a whole? If this is the case then it casts doubt on whether the first of the four properties of the behaviors of parts within a functioning system (i.e. being flourishing enhancing or maintaining) is really a property of systems.

This objection does not only cast doubt on the first property of good behaviors of parts. It also brings into question the two rules that are generated by social systems with parts that have control over their own behaviors. Given the centrality of these rules to morality, the fact that the analogy appears to generate them is important. One unpromising line of response is to dismiss the objection as citing an anomalous system. After all, most systems do not have as one of their purposes to directly destroy themselves. However, pursuing this option would be a mistake. Though most systems do not include a self-destruct mechanism, there are many other systems that will lack the first property discussed above. It will be helpful here to look at the difference between mechanical systems and human social systems. A mechanical system is designed to serve the purposes of its designer. Frequently, the designer's purposes will be best served by the continued functioning of the system but this will not always be the case. In addition to a self-destruct mechanism, it is sometimes in the interests of businesses to build a

sub-optimal system so as to maximize profit. For example, a car whose engine and transmission will only last for 100,000 miles before it needs to be rebuilt will force its owner to either do costly repairs or to buy a new car much sooner than a car whose engine and transmission are designed to last longer. This will serve the interests of those who can profit from the sale of new cars if it is likely enough that the person will buy another car from that company. Further, if all the companies are making their cars like this then it can serve the interests of the car manufacturers even if people will buy their next car from a different manufacturer than the last one.

Fortunately, we can restrict the analogy so as to exclude these systems, which I will call short-lived systems, and do so in a way that is not ad hoc. We can see this by considering an important difference between mechanical systems and human social systems. The existence of the latter is largely a result of natural selection. Just how these systems have been selected for is a large question but much of this work has already been done. It is evident that altruistic behavior between the parts within these social systems is essential for the continued functioning of the systems. If we were willing to exploit a neighbor whenever we saw an apparent advantage in it, we would wind up with something like Thomas Hobbes' war of all against all. However, given that cooperation between members of non-human species including among chimpanzees and bonobos tends to be quite common, it is plausible to think human beings never really lived in such an environment.

One important element of the altruistic behavior we see among organisms within the natural world comes from what is called inclusive fitness. A key to understanding how this type of helping behavior can come about involves recognizing that natural selection occurs primarily at the level of genes. As a result, helping behavior between related individuals can be selected for even when the helping behavior involves a fitness cost to the helper himself. The reason is that kin share many of the same genes. If an organism 'A' helps out one of its kin 'B' then even if the helping behavior has a fitness cost for A, the benefit to B may be great enough that it

offsets the cost to A, at least from the perspective of the genes that get reproduced. Further, if organisms that have these genes do better than ones that do not (because they are likely to get help from kin who also have the genes) then we can expect these genes to eventually come to dominate in future generations of the population.

Occasionally human systems will arise which destroy themselves. Sometimes cults will come about in which the members commit mass suicide but these are an anomaly. We recognize that something has gone wrong when this occurs. Human social systems tend to be long-lived rather than short-lived systems. It is reasonable to postulate that the reason is that the fate of our genes is largely tied up with the fate of the bodies that house them. Occasionally sacrificing for kin will lead to the destruction of our bodies and the genes that inhabit them but usually they will not. Plus, we can usually count on our kin to help us out as well. Notice also that when there is marked instability in human social systems that there is usually a high price paid for people, as well as genes, that inhabit those social systems. Revolutions are often bloody. For these reasons, we can safely eliminate short-lived systems from consideration.

#### Objection 2 and response

A second objection stems from the claim that the two general moral rules—*Do not harm others* and *Help others*—are generated from the nature of moral systems. Is it specifically moral systems that generate these rules and exactly what is meant by generation?

In responding to the previous objection we made a distinction between short-lived and long-lived systems and restricted our attention to long-lived systems. This move was justified because evolution would select for long-lived social systems rather than short-lived ones due to the fact that the former would help promote the interests of the genes whereas the latter kinds of systems would tend to undermine those interests. We need to make a further restriction in the systems we are considering. We need to restrict the type of system to ones the behavior of whose parts exhibits a good deal of plasticity. Thinking about this helps us to see a further

difference between human social systems and mechanical systems as well as how human social systems differ from those that develop among animals like bees and ants. The behaviors of insects is quite rigidly programmed and is so to such an extent that it probably does not make sense to say that an ant or bee chooses to do one thing rather than another. For example, a digger wasp will sting and paralyze an insect and bring it back to its nest.<sup>274</sup> The digger wasp positions its prey just outside the entrance to the nest and goes in to inspect the nest before bringing in the prey. If the prey is moved a few inches from the nest entrance while the wasp is doing its inspection the wasp will move it back to the entrance and then go in and inspect again. The wasp's behavior seems to be rigidly programmed; it places the prey at the entrance to the nest, it goes in to inspect and then it drags its prey in. Even though it has just inspected the nest, if the prey is moved it has to restart the sequence.

Returning to the automotive example, though a cooling fan is not a living being it has no more ability to vary its behavior than does the digger wasp. Human beings, on the other hand, exhibit a great deal of plasticity in their behaviors. Our environments have a great deal of influence over the behavioral tendencies that we acquire. Sometimes this is because others approve of a certain type of behavior. Other times it is because we have figured out a way to solve certain problem.<sup>275</sup> If we were rigidly programmed like the digger wasp or the cooling fan, we would have no need for moral rules, not that we could choose to follow them anyway.

The plasticity in human behavior opens up the possibility that we will behave in ways that undermine the system of which we are a part. Some of the behaviors we adopt may be detrimental to the overall functioning of the system. One of the most fundamental ways we can do this is by doing damage to other parts of the system. Further, if we can help other parts of the system that are not flourishing this will tend to have a positive effect on the functioning of the system. Because we are not rigidly programmed, the two general rules will be necessary

---

<sup>274</sup> This example is discussed in Dennett 1984.

<sup>275</sup> I am not implying that all behaviors are adaptive.

components of long-lived human social systems. Let me be specific about what I mean by necessary components. People will have to exhibit enough of a propensity to behave as if they were following these rules so as not to undermine the system. To some extent, this behavior appears to have roots in our biology. For example, babies seem to exhibit some ability to empathize with each other. A baby who is in the presence of a distressed baby will likely become distressed him or herself. This tendency could work to inhibit aggressive behavior and prompt helping behavior to relieve the distress. Yet, we know all too well that these kinds of mechanisms are not sufficient to ensure that human beings follow the two rules. Working alongside mechanisms that lead to moral behavior are ones that tend to lead to immoral behavior. If we can benefit ourselves, as well as our genes, by behaving in immoral ways then the forces of natural selection will presumably select for these behaviors as well. The plasticity of human behavior coupled with the inability of behavioral mechanisms to ensure we do not undermine the system necessitates the need for the two general rules to be a part of our moral system. As such, these rules should be found in some sense or another in all human cultures. Sometimes people outside the social group will be seen as not covered by these rules. Frequently, the rules may be stated in more specific terms (you must not push your sister down, you ought to help out the less fortunate, etc.) but the general rules should be at work within every functioning human society.

It would be a mistake to interpret my claims about human behavior as a claim that human beings possess libertarian-style free will. Even if determinism is true, there are a vast number of factors that affect human decisions and a great many of these depend on what the individual has learned. This makes it unlikely that we could ever learn enough rules to cover every possible situation in which we might find ourselves. The less rigidly programmed the parts of the system, whether by a conscious designer as with a mechanical system or by the blind forces of evolution as is the case with human beings, the more there will be a need for general rules to prevent the parts from destroying the system in which they are ensconced.



### Objection 3 and response

The discussion of natural selection leads to a third objection. The appeal to mechanisms of natural selection, like inclusive fitness, makes the mechanical systems analogy implausible. The reason is that there is no way to reduce morality to the working of such mechanisms. Let's think about helping kin. If helping kin helps me to get my genes passed down to the next generation and helping non-kin does not do so then it appears to follow that I ought not to help non-kin. If I pass a pool in which a child is drowning and I know the child is not related to me it seems that I would have no moral obligation to help the child. This may seem to lead to a straightforward reductio of the view.

Further the appeal to additional biological mechanisms will not provide a satisfying fix for the general problem. For example, we might discuss other theories of biological mechanisms thought to be relevant to moral behavior and calculate that when the effects of these mechanisms are considered together they indicate that I ought to save the drowning child. Yet skepticism about this approach is warranted. We seem to be required to take it on faith that we will get the answers that we think we ought to get. Yet the forces of natural selection do not care about morality. More importantly, we would need some non-arbitrary way to distinguish between which mechanisms lead to moral behavior and which do not. Consider the following example. Human males sometimes appear to engage in a behavior called mate guarding. The purpose of mate guarding is to prevent one's female partner from having the opportunity to reproduce with anyone else. For the perspective of the genes, it makes sense to engage in this behavior because males who invest resources in raising children who are not related to them are making it less likely that their genes will get passed down to subsequent generations. Different cultures have different beliefs concerning the moral status of such behaviors. In some western cultures, such behaviors are often viewed as a violation of the woman's rights,

especially if they lead to violence or a limitation on the woman's autonomy. In some other cultures, women are not allowed to leave the home without a male escort. In fact, the practice of female genital cutting, which sometimes involves the removal of the entire clitoris, can be seen as particularly disturbing form of mate guarding. Further, these behaviors are often viewed as being morally right. On what basis are we to say that engaging in mate guarding behaviors are not right or even wrong if evolution has selected for this trait?

Fortunately, the view on offer here can avoid these serious objections. The analogy does not entail that morality is a product of certain evolved behaviors such as inclusive fitness, reciprocal altruism, or mate guarding (or a combination of the three). The view states that moral properties arise from human social systems themselves. The question of which forces have made us such that we tend to form such systems is an interesting question in its own right but need not be answered here. Appealing to an etiological theory of functions may provide an explanation of why we have evolved to form long-lived social systems but it will not provide us with a straightforward explanation of right and wrong. The important point is that human beings do tend to form functioning social systems and that the main aspects of morality can be determined by looking at the properties of those systems themselves. With this in mind, I return to a defense of the properties of those systems.

#### Objection 4 and response

In considering the first two objections I argued that the first property of the good behavior of parts within a human social system and the two general moral rules that arise from human social systems are properties of long-lived systems but are not necessary properties of all systems. Further, I argued that human social systems just are these kinds of systems because the beings that create them are a product of natural selection. For the mechanical systems analogy to be successful it has to explain enough of the main aspects of morality and it cannot get too much wrong. Though lacking in precision, the preceding statement is essentially what

the method of reflective equilibrium advises in our decisions to accept or reject a moral principle or theory. With that in mind, I consider a fourth objection.

The third aspect of behaviors with moral relevance within a functioning system is that such behaviors were ones that affected other parts of the system. This says nothing about behaviors which only affect the one performing the behavior. Some might object that there are realms of behavior that do not affect others but which still have moral relevance. Suicide, drug use and so-called deviant sexual behaviors are prime examples. If the mechanical systems analogy delivers the wrong answers on these questions then this gives us reason to doubt the relevance of the analogy.

First, notice that the moral status of each of these is open to question in a way that some behaviors that affect others are not. We do not seriously question whether killing innocent people for the fun of it, theft and rape are morally permissible but many of us think that suicide, recreational drug use and masturbation are. Few people think masturbation is actually morally wrong but note that even for the more controversial cases of suicide and recreational drug use, the wrongness that comes from such supposedly self-affecting behaviors may come from the effects that such behaviors are thought to have on other people. J. S. Mill makes this point in the following passage of *On Liberty*.

I fully admit that the mischief which a person does to himself may seriously affect, both through their sympathies and their interests, those nearly connected with him and, in a minor degree, society at large. When, by conduct of this sort, a person is led to violate a distinct and assignable obligation to any other person or persons, the case is taken out of the self-regarding class and becomes amenable to moral disapprobation in the proper sense of the term.<sup>276</sup>

Mill goes on to give examples of a man who irresponsibly incurs debts he cannot repay and a policeman who is drunk on duty. Given that these behaviors that affect ourselves and cause self-harm are likely to harm others as well, they are really ought to be thought of as other affecting. As a result, the realm of behaviors which are not other-affecting is probably smaller

---

<sup>276</sup> Mill 1978, p. 79.

than is commonly thought. What the mechanical systems analogy reveals is that the wrongness of these behaviors, when they are wrong, comes from the affects they have on others and not the affects they have on the individuals themselves.

The objector may continue that there are cases in which a behavior affects no one else and is still morally wrong. Two examples come to mind. First, we can imagine a hermit who commits suicide. His behavior affects no one else but it is still wrong. In reply to this first example, my intuition is that our judgment ought to be that the hermit's behavior is unfortunate. We ought to be sad that he was driven to such a behavior but that, strictly speaking, he has done nothing morally wrong. Further, we might be tempted to judge his behavior as morally wrong because of our judgments in similar cases in which others are affected. If suicide is wrong under most circumstances then we are tempted to say that it is wrong in this case as well. If the objector persists in her claim that the hermit's behavior is morally wrong then, according to the view we are discussing here, she is wrong.

Second, a person may insist that masturbation is morally wrong even though it only affects the person committing the act and does not make her less able to function in society. According to the view being defended here, the objector is wrong. It is important to remember that a moral theory does not need to validate all pre-theoretic intuitions that anyone has but only those that remain robust under scrutiny.<sup>277</sup>

#### Objection 5 and response

The fifth objection concerns the nature of political systems. Observations of the world in which we live seem to indicate that there is a significant variety in existing political systems.

---

<sup>277</sup> One cannot help but be reminded of the four examples that Kant gives of applications of the categorical imperative in chapter two of the *Groundwork of the Metaphysics of Morals*. His examples regarding duties to others—the duty to refrain from making false promises and the duty to help those in need—seem much more plausible than his duties to self—the duty to refrain from committing suicide and the duty to develop one's talents. In my experience, when students defend Kant's claims on the latter duties they often cite (though wrongly from Kant's perspective) the way others will be effected when these duties are shirked.

Further, some of the political structures appear to be explicable by appealing to natural selection. For example, it is not uncommon to see positions of power awarded to family members where this is allowed. In many monarchies and dictatorships, power passes from a parent to a child, usually from father to son. Inclusive fitness seems to provide an explanation for why we would do this. The ruler has more power than others and presumably will have more access to food, shelter, healthcare, etc. As a result, any offspring they produce are more likely to be able to pass their genes down to future generations. Peter Railton argues that unjust social systems will lead to discontent which may in turn lead to more just arrangements. However, he also notes that other forces will be at work that may have the opposite effect.<sup>278</sup> As was discussed in response to objection 3, we can expect that natural selection taking place at the level of genes will tend to lead to forces that work both for and against what we would typically think of as moral behavior. Further, an efficiently run, oppressive but not too brutal dictatorship might be a perfectly good political system from the point of view of our genes as long as its subjects are able to successfully reproduce. In fact, the stability that comes from such a regime might even be a plus. With these considerations in mind, why should we prefer Aristotle's ideal to any other?

There is reason to think that questions in political philosophy are primarily questions in applied moral philosophy even though political philosophy and moral philosophy are frequently treated as different subjects. Just as we may ask questions about the moral status of the practice of abortion or stem cell research we may ask whether a democracy or dictatorship is morally preferable. In *On Liberty*, J. S. Mill gives the following explanation of his project.

I regard utility as the ultimate appeal on all ethical questions; but it must be utility in the largest sense, grounded on the permanent interests of man as a progressive being. Those interests, I contend, authorize the subjection of individual spontaneity to external control only in respect to those actions of each which concern the interest of other people.<sup>279</sup>

---

<sup>278</sup> Railton 1984.

<sup>279</sup> Mill 1978, p. 10

Mill is trying to answer a question about how a political system ought to be run and recognizes that this is a question in applied ethics. He argues that we ought not to interfere with the rights of an individual when that person's behavior only affects him or herself and that we can only interfere to protect ourselves from harm that they would do us. Mill's justification for this is grounded in his view of morality. Adopting Mill's *harm-only* principle would maximize utility within society. Numerous other examples of philosophers treating political philosophy as an extension of moral considerations to political or social institutions can be found but I will not discuss them here.

As is probably apparent from our earlier discussion, to ask what is the most moral or just state is not to ask about the function of states. Morality arises from the nature of human social systems themselves; the cause of those systems is beside the point. Let us assume that the causal explanation for why human beings are the kinds of organisms that form social systems is that such social systems served the interests of the genes that helped to create them. However, having arisen, they give birth to morality. The four properties of good behaviors of the parts hold for any parts within a long-lived system regardless of the cognitive capacities of its parts. However, when the parts reach a level of cognitive sophistication such that they can adapt their behavior to rules and have enough plasticity in their behavior such that they can act to the detriment of the social system of which they are a part, the two general rules will apply.<sup>280</sup> Yet there is no guarantee that the state that most effectively realizes the purposes for which states (i.e. social systems) evolved will be the most morally just state. Alterations in states to make them more morally may or may not make it more likely that their members will be able to pass down their genes to subsequent generations.

These realizations will lead us to prefer an Aristotelian view of the *morally* ideal social system even if it is not, strictly speaking, Aristotle's view. One way in which the considerations

---

<sup>280</sup> The fact that these rules will be chosen does not entail that we would not have evolved some tendency to act as they would dictate.

discussed here lead us to a view that differs from Aristotle's is that he seems to have a view of a rational person deciding to form a state because doing so is in her or his best interests. As such, it makes sense that the goodness of the state will be reducible to the goodness of its parts and the purpose of the state will be reducible to the purposes of those who formed the state. Yet, it is implausible to think that human involvement in the complex social systems is something we would usually call a choice. Given that some of our closest evolutionary "relatives", namely chimpanzees and bonobos, form complex social systems it is unlikely that we ever chose this. The purpose of these systems is probably better thought of as reducible to the purposes of the genes that get reproduced.

Yet the mechanical systems analogy provides some insight into why it is probably most accurate to think of moral goodness of a human social system as reducible in some important sense to the flourishing of its parts. An example will help to illustrate this. To cite a familiar and disturbing example, some societies require what is sometimes called female circumcision in which the entire clitoris is removed. This procedure may be performed in unsanitary conditions with little or no anesthesia. The view being offered here might at first seem to dictate that if such a practice is required in a particular culture it is actually good there. The first property of the good behaviors of parts within a long-lived system is that they maintain or enhance the flourishing of the system itself. This may seem to license a kind of conservatism in which whatever practices exist within the social system are deemed good.

However, the existence of the two general rules which emerge in human social systems can lead to the revision of moral systems in such a way that, in practice, the goodness of a moral system is reducible to the flourishing of its parts. If there is a practice which causes harm to a person or group of people and it is not necessary to prevent harm to a different group of people then the two rules speak against it. Given that the particular institutions and practices within social systems vary considerably and that many of these practices are merely contingent

on the particular history of that social group, the practices ought to be changed when they violate the two rules.

### 5.8 Summary

This chapter defends a naturalistic solution to the moral problem without sacrificing moral objectivity. It was argued that there are at least three theories of mental states that can explain both the descriptive and the directive nature that moral mental states appear to have much of the time. Further, if we focus on a virtue-based account of morality then it will not be necessary for moral agents to be tracking moral properties but only morally relevant properties. The virtue ethicist can defend the objectivity of a moral claim by citing what the virtuous person would do in a given situation. The virtuous person's actions are justified because they are necessary in order for their possessor to live a flourishing life.

The virtue ethicist can respond to the objection that flourishing is possible without possessing the virtues by appealing to the larger social structure that Aristotle thought was essential to a full explanation of how one ought to live. By drawing an analogy with mechanical systems, we can make clear why Aristotle's virtues and a political structure which promotes them are morally preferable to other character trait/political structure pairings, even if some people can flourish in other social structures. First, four properties of a correctly functioning part within a mechanical system were highlighted in order to show the similarity between a mechanical system and a moral system. Second, we discussed some of the aspects of the ideal state identified by Aristotle. It was argued that much of what he is trying to do after settling on what the goals of the ideal state will be is describing ways to protect and promote the goals of the ideal state. Returning to the mechanical systems analogy, we saw that there were two general rules or principles that arose from the analogy and that these could be used as general guidelines for structuring the society. Developing the analogy revealed a non-arbitrary method for choosing a proper subset of the possible character trait/political structure sets. Further, the



analogy supports a plausible semantic analysis that respects our metaphysical intuitions on thin moral terms. There is further work yet to be done and I have indicated some ways this might be pursued within this chapter.

## Bibliography

- Altham, J. E. J. 1986: "The Legacy of Emotivism." In McDonald, G. and Wright, C. 1986.
- Aristotle. 2001 (1940). *The Basic Works of Aristotle*. Edited by Richard McKeon. New York: The Modern Library.
- Aristotle. *De Anima*. Translated by J. A. Smith. In Aristotle 2001 (1940).
- Aristotle. 1999. *Nicomachean Ethics*. Translated by Terence Irwin. Indianapolis and Cambridge: Hackett Publishing Company.
- Aristotle. *Politics*. Translated by Benjamin Jowett. In Aristotle 2001 (1940).
- Axelrod, Robert. 1985. *The Evolution of Cooperation*. Basic Books.
- Ayer, Alfred. J. 1952 (1946): *Language, Truth and Logic*. 2<sup>nd</sup> Edition. Dover Publications, Inc: New York.
- Blackburn, Simon. 1993: *Essays in Quasi-Realism*. New York: Oxford University Press.
- Borg, Emma. 2004. *Minimal Semantics*. New York: Oxford University Press.
- Brink, David O. 1986: "Externalist Moral Realism." *The Southern Journal of Philosophy*, Vol. XXIV, Supplement, 23-41.
- Brink, David O. 1989: *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Brink, David O. 1997: "Moral Motivation." *Ethics*, Vol. 108, No. 1, pp. 4-32.
- Burnyeat, M. F. 1980. "Aristotle on learning to be good." In Sherman 1999.
- Butler, R. J. (ed) 1965: *Analytical Philosophy: Second Series*. New York: Barnes and Noble.
- Cappelen, Herman and Lepore, Ernie. 2005. *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Malden: Blackwell Publishing.
- Churchland, Paul M. 1981: "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78, No. 2, 67-90.
- Conly, Sarah. 1988. "Flourishing and the Failure of the Ethics of Virtue." In French, Uehling, Jr. and Wettstein (eds.).
- Copp, David. 1995: *Morality, Normativity and Society*. New York: Oxford University Press.
- Copp, David (ed.). 2006: *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
- Copp, David. 2008: "Darwinian Skepticism About Moral Realism." *Philosophical Issues*, (18), *Interdisciplinary Core Philosophy*, 186-206.

Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *The Journal of Philosophy*, Vol. 76, No. 5, 256-282.

De Lancey, Craig. 2002. *Passionate Engines: What Emotions Reveal About Mind and Artificial Intelligence*. Oxford: Oxford University Press.

Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.

De Vries, Willem A. and Triplett, Timm. 2000. *Knowledge, Mind, and the Given: Reading Wilfred Sellars's "Empiricism and the Philosophy of Mind," including the complete text of Sellars's essay*. Indianapolis/Cambridge: Hackett Publishing Company, Inc.

Doris, John. 1998. "Persons, Situations, and Virtue Ethics." *Noûs*, 32, (4), 504-530.

Doris, John. 2002. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Dreier, James. 2000. "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research*, Vol. 61, No. 3, 619-638.

Fodor, Jerry A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge: The MIT Press.

Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review*, Vol. 81, No. 3, 305-316.

Foot, Philippa. 1977. "Approval and Disapproval." In Foot 2002.

Foot, Philippa. 1978. "Are Moral Considerations Overriding?" In Foot 2002.

Foot, Philippa. 2002. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Clarendon Press.

French, Peter A., Uehling, Jr., Theodore E. and Wettstein, Howard K. (eds.) 1988. *Midwest Studies in Philosophy Volume XIII Ethical Theory: Character and Virtue*. Notre Dame: University of Notre Dame Press.

Gendler, Tamar Szabó. 2008a. "Alief and Belief." *The Journal of Philosophy*, 634-663.

Gendler, Tamar Szabó. 2008b. "Alief in Action (and Reaction)." *Mind & Language*, Vol. 23, No. 5, 552-585.

Gendler, Tamar Szabó. 2012. "Between Reason and Reflex: Response to Commentators." *Analysis*, 72, (4), 799-811.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Hamilton, W. D. 1964. "The Genetical Evolution of Human Behavior." *Journal of Theoretical Biology*, 7, (1), 1-16.

- Hansell, Mike. 2000. *Bird Nests and Construction Behavior*. Cambridge: Cambridge University Press.
- Hardie W. F. R. 1980. *Aristotle's Ethical Theory*, 2<sup>nd</sup> Edition. Oxford: Clarendon Press.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Clarendon Press.
- Hare, R. M. 1963. *Freedom and Reason*. Oxford: Clarendon Press.
- Hare, R. M. 1981. *Moral Thinking*. Oxford: Clarendon Press.
- Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
- Hume, David. 1978. *A Treatise of Human Nature*, 2<sup>nd</sup> edition. Oxford: Oxford University Press.
- Hursthouse, Lawrence, and Quinn. 1995. *Virtues and Reasons-Philippa Foot and Moral Theory: Essays in Honor of Philippa Foot*. Oxford: Clarendon Press.
- Jackson, Frank. 1998: *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.
- Joyce, Richard. 2001: *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, Richard. 2006: *The Evolution of Morality*. Cambridge: MIT Press.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kant, Immanuel. 1781. *Metaphysical Foundations of Morals*. In Kant 1993.
- Kant, Immanuel. 1993. *The Philosophy of Kant: Immanuel Kant's Moral and Political Philosophy*. New York: The Modern Library.
- Kappel, K. 2006. "The Meta-Justification of Reflective Equilibrium." *Ethical Theory and Moral Practice*, 9, 131-147.
- Kelly, Thomas and McGrath, Sarah. 2010: "Is Reflective Equilibrium Enough?" *Philosophical Perspectives*, 24, 325-359.
- Keren, Gideon and Schul, Yaacov. 2009. "Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories." *Perspectives on Psychological Science*, 4, 533-550.
- Kitcher, Philip. "Biology and Ethics." In Copp 2006.
- Kraut, R. "Aristotle on Becoming Good: Habituation, Reflection, and Perception." In *The Oxford handbook of Aristotle*. Edited by ? Oxford and New York: Oxford University Press.
- Kristjansson, Kristjan. 2012. "Aristotelian motivational externalism." *Philosophical Studies*, published online: 31 January 2012.
- Lillehammer, Hallvard. 1997. "Smith on Moral Fetishism." *Analysis*, 57.3, 187-195.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.

- Mason, Elinor. 2008. "An Argument Against Motivational Internalism." *Proceedings of the Aristotelian Society*, Vol CVIII, Part 2, 135-156.
- Matthews, Gareth B. 1986. "Gender and Essence in Aristotle." *Australasian Journal of Philosophy*, Supplement to Vol. 64, 16-25.
- McDonald, G. and Wright, C. 1986. *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*. Oxford: Basil Blackwell.
- McDowell, John. 1978. "Are Moral Requirements Hypothetical Imperatives?" *The Aristotelian Society*, Supplementary Volume LII, 13-29.
- McDowell, John. 1979. "Virtue and Reason." *Monist: An International Quarterly Journal of General Philosophical Inquiry*, 62, 331-350.
- McDowell, John. 1995. "Two Sorts of Naturalism." In Hursthouse, Lawrence, and Quinn.
- McKeon, Richard. 2001 (1941). *The Basic Works of Aristotle*. New York: The Modern Library.
- McNaughton, David. 1988. *Moral Vision: An Introduction to Ethics*. Oxford: Blackwell Publishers, Inc.
- Mill, J. S. 1978 (1859). *On Liberty*. Indianapolis: Hackett Publishing Company, Inc.
- Miller, Alexander. 1996. "An objection to Smith's argument for internalism." *Analysis*, 56.3, 169-174.
- Miller, Jr., Fred D. 1995. *Nature, Justice, and Rights in Aristotle's 'Politics'*. Oxford: Clarendon Press.
- Millikan, Ruth Garrett. 1995. "Pushmi-Pullyu Representations." *Philosophical Perspectives*, Vol. 9, *AI, Connectionism and Philosophical Psychology*, 185-200.
- Millikan, Ruth. 2004. *Varieties of Meaning*. Cambridge, Ma., and London: MIT Press.
- Nussbaum, Martha C. 1992. "Human Functioning and Social Justice: In Defense of Aristotelian Essentialism." *Political Theory*, Vol. 20, No. 4, 202-246.
- Nussbaum, Martha C. and Rorty, Amelie O. 1992. *Essays on Aristotle's De Anima*. Oxford: Clarendon Press.
- Olson, Jonas. 2002. "Are Desires De Dicto Fetishistic?" *Inquiry: An Interdisciplinary Journal of Philosophy*, 45:1, 89-96.
- Putnam, Hilary. 1963: "Brains and Behavior." In R. J. Butler (ed.)
- Railton, Peter. 1984. "Moral Realism." *The Philosophical Review*, XCV, No. 2, 163-207.
- Rawls, John. 1971 (1999). *A Theory of Justice*. Revised edition. Cambridge, Ma: The Belknap Press of Harvard University Press.

- Rawls, John. 1974-1975. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association*, Vol. 48, 5-22.
- Richardson, Henry S. 1992. "Desire and the Good in *De Anima*." In Nussbaum and Rorty, (eds.).
- Sadler, Brook J. 2003. "The Possibility of Amoralism: A Defence Against Internalism." *Philosophy*, Vol. 78, No. 303, 63-78.
- Sayre-McCord, Geoffrey (ed). 1988. *Essays on Moral Realism*. Ithaca: Cornell University Press.
- Scheffler, Samuel. 2011. "Valuing." In Wallace, R. J., Kumar, R., and Freeman, S. (eds.) 2011
- Sellars, Wilfred. 1956. "Empiricism and the philosophy of mind." In DeVries and Triplett.
- Shafer-Landau, Russ. 1998. "Moral Judgment and Moral Motivation." *Philosophical Quarterly*, Vol. 48, No. 192, 353-358.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Sherman, Nancy. 1989. "The Habituation of Character." In Sherman 1999.
- Sherman, Nancy. 1999. *Aristotle's Ethics: Critical Essays*. Lanham: Rowman & Littlefield Publishers, Inc.
- Smith, Michael. 1988. "On Humeans, Anti-Humeans and Motivation: A Reply to Pettit." In Smith 2004.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Basil Blackwell.
- Smith, Michael. 2004. *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-ethics*. Cambridge: Cambridge University Press.
- Smith, Michael. 2007. "In Defence of *Ethics and the A Priori*: A Reply to Enoch, Hieronymi, and Tannenbaum." *Philosophical Books*, Vol. 48, No. 2, 136-149.
- Sobel, David. 1999. "Do the desires of rational agents converge?" *Analysis*, 59, 3, 137-147.
- Sober and Wilson. 1998. *Unto Others*. Cambridge, Mass: Harvard University Press.
- Stampe, Dennis W. 1986. "Defining Desire." From Joel Marks (ed.) *The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting*. Chicago: Precedent Publishing, Inc., 149-173.
- Stampe, Dennis W. 1987. "The Authority of Desire." *The Philosophical Review*, XCVI, No. 3, 335-381.

- Stampe, Dennis W. 1990. "Desires as Reasons—Discussion Notes on Fred Dretske's 'Explaining Behavior: Reasons in a World of Causes'." *Philosophy and Phenomenological Research*, Vol 50, No. 4, 787-793.
- Stocker, Michael. 1979. "Desiring the Bad: An Essay in Moral Philosophy." *The Journal of Philosophy*, Vol. 76, No. 12, 738-753.
- Stocker, Michael. 1981. "Values and Purposes: The Limits of Teleology and the Ends of Friendship." *The Journal of Philosophy*, Vol. 78, No. 12, 747-765.
- Stout, Rowland. 1996. *Things That Happen Because They Should: A Teleological Approach to Action*. Oxford and New York: Oxford University Press.
- Strandberg, Caj. 2000. "Aristotle's internalism in the *Nichomachean Ethics*." *Journal of Ethics*, 34 (1), 71-87.
- Strandberg, Caj. 2011. "The Pragmatics of Moral Motivation." *Journal of Ethics*, 15, 341-369.
- Strandberg, Caj. 2012. "A Dual Aspect Account of Moral Language." *Philosophy and Phenomenological Research*, LXXXIV, No. 1, 87-122.
- Strandberg, Caj. 2013. "An Internalist Dilemma—and an Externalist Solution." *Journal of Moral Philosophy*, 10, 25-51.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies*, 127, 109-166.
- Street, Sharon. 2008. "Reply to Copp: Naturalism, Normativity and the Varieties of Realism Worth Worrying About." *Philosophical Issues*, Vol. 18, (1), 207-228.
- Sturgeon, Nicholas. 1984. "Moral Explanations." In Sayre-McCord (ed.) 1988.
- Sullivan, Shannon. 2006. *Revealing whiteness: the unconscious habits of racial privilege*. Bloomington: Indiana University Press.
- Svavarsdottir, Sigrun. 1999: "Moral Cognitivism and Motivation." *The Philosophical Review*, Vol. 108, No. 2, 161-219.
- Thomson, Judith Jarvis. 2008. *Normativity*. Chicago and LaSalle: Open Court.
- Toppinen, Teemu. 2004. "Moral Fetishism Revisited." *Proceedings of the Aristotelian Society*, 104, (3), 305-313.
- Trivers, R. L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, 46, 35-57.
- Wallace, R. J., Kumar, R., and Freeman, S. 2011. *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford and New York: Oxford University Press.
- Wheeler III, Samuel C. 1972. "Attributives and their Modifiers." *Noûs*, Vol. 6, No. 4, pp. 310-334.

Williams, Bernard. "Internal and external reasons." In Williams 1981.

Williams, Bernard. 1976: "Persons, character and morality." In Williams 1981.

Williams, Bernard. 1981: *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.

Witt, Charlotte. 2011. *The Metaphysics of Gender*. Oxford: Oxford University Press.

Wong, David B. "Emotion and the Cognition of Reasons in Moral Motivation." *Philosophical Issues*, 19, *Metaethics*, 343-367.

Zangwill, Nick. 2008a: "Desires and the Motivational Debate." *Theoria*, 74, 50-59.

Zangwill, Nick. 2008b: "The indifference argument." *Philosophical Studies*, 138, 91-124.

Zangwill, Nick. 2010: "Normativity and the Metaphysics of Mind." *Australasian Journal of Philosophy*, Vol. 88, No. 1, pp. 21-39.