MEETING ANSCOMBE’S DEMAND:
TOWARD A MORAL PSYCHOLOGY OF CHARACTER

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Liam Murphy Monahan, B.A., M.A.

__________________________________________
W. David Solomon, Director

Graduate Program in Philosophy
Notre Dame, Indiana

May 2004
MEETING ANSCOMBE’S DEMAND:
TOWARD A MORAL PSYCHOLOGY OF CHARACTER

Abstract

by

Liam Murphy Monahan

Contemporary virtue theorists suggest a conception of moral perception, practical reasoning, and motivation radically different from that of their Enlightenment-inspired rivals. However, without a psychologically credible understanding of how it is possible for rationally limited creatures like human beings to be virtuous, virtue theory cannot be a genuine alternative. Many virtue theorists believe that being virtuous involves the possession of stable character traits that provide an agent with the capacity to consistently recognize objectively appropriate reasons for action as well as the ability to act reliably for, or because of, those reasons. What many virtue theorists fail to realize is that this description of being virtuous makes both normative and psychological assumptions that must be vindicated.

I argue that no normative account of the existence of objectively appropriate reasons for action can be derived from an agent-based account of right action. An
agent-based account of right action claims that evaluations of human action are derived from the motivation or behavior of a virtuous agent. I conclude that in order for virtue ethics to be a viable alternative moral theory, it must begin with a substantive account of the good.

I also argue that a theory concerning the moral psychology of a virtuous agent can be constructed from Michael Bratman’s descriptive account of his temporally extended planning theory of human agency. By identifying the primary functional roles played by the virtues in the psychology of a virtuous agent, I conclude that an agent’s commitments can become embedded in her intentions, plans and policies in such a way that her commitments take on a central role in: (1) directing her attention to what is morally relevant, (2) framing her practical reasoning around the good, and (3) motivating her to pursue appropriate ends. When this account of the moral psychology of a virtuous agent is coupled with an adequate substantive account of the good, virtue theory stands a chance of becoming a genuine rival moral theory.
To my wife, Amy Monahan, and my parents, Shawne and Michael Monahan, with love and gratitude
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... v

CHAPTER 1: ANSCOMBE’S DEMAND .................................................................................. 1
  1.1 Introduction .................................................................................................................. 1
  1.2 The Normative Condition .......................................................................................... 2
  1.3 The Psychological Condition ..................................................................................... 5
    1.3.1 The Functional Roles of Virtues .......................................................................... 6
    1.3.2 Anscombe’s Demand ......................................................................................... 12
    1.3.3 Failures to Meet Anscombe’s Demand .............................................................. 13
    1.3.4 A Sketch of a Strategy to Meet Anscombe’s Demand ........................................ 16
  1.4 A Wittgensteinian Objection ....................................................................................... 19
  1.5 Conclusion .................................................................................................................. 28

CHAPTER 2: THE NORMATIVE CONDITION OF POSSIBILITY FOR BEING VIRTUOUS ................................................................ ..................................................... 30
  2.1 Introduction .................................................................................................................. 30
  2.2 Reasons for Action ...................................................................................................... 33
    2.2.1 Objections and Replies ...................................................................................... 44
  2.3 The Rejection of Non-teleological Agent-Based Virtue Theories ................................ 47
    2.3.1 Objections and Replies ...................................................................................... 59
  2.4 The Rejection of Teleological Agent-based Virtue Ethics .......................................... 60
    2.4.1 Hursthouse as a Teleological Agent-based Virtue Theorist ................................. 63
    2.4.2 A Teleological Good-based Virtue Theory ...................................................... 69
    2.4.3 Replies and Objections ..................................................................................... 72
  2.5 The Structure of Normative Theory ............................................................................. 77
  2.6 Conclusion .................................................................................................................. 79

CHAPTER 3: THE TEMPORALLY EXTENDED PLANNING THEORY OF HUMAN AGENCY ......................................................................................................................... 81
  3.1 Introduction .................................................................................................................. 81
  3.2 Intentions, Plans, and Commitment ............................................................................ 84
  3.3 Commitment ................................................................................................................ 93
  3.4 The Temporally Extended Planning Theory of Human Agency .............................. 101
    3.4.1 Rejecting the Desire-Belief Model of Human Agency ....................................... 103
      3.4.1.1 Reductive Conception of Intention: Version 1 ............................................ 104
      3.4.1.2 Reductive Conception of Intention: Version 2 .......................................... 106
    3.4.2 Description of the Temporally Extended Planning Theory of Human Agency .......... 107
  3.5 Conclusion .................................................................................................................. 133
CHAPTER 4:  THE PSYCHOLOGICAL CONDITIONS OF POSSIBILITY FOR BEING VIRTUOUS ................................................................. 134
  4.1 Introduction .................................................................................................. 134
  4.2 A Good-based Foundational Account of Reasons for Action ..................... 136
  4.3 Toward a Moral Psychology of Character .................................................. 152
  4.4 Having Character and Being Virtuous ....................................................... 161
  4.5 Fulfilling the Functional Roles of the Virtues ............................................. 168
    4.5.1 Operate in Human Agency ................................................................. 168
    4.5.2 Stability and Entrenchment ............................................................... 171
    4.5.3 Maintain Moral Harmony ................................................................. 173
    4.5.4 Narrative Unity .................................................................................. 176
    4.5.5 Constitute Eudaimonia ...................................................................... 178
    4.5.6 Voluntary Acquisition and Modification ........................................... 181
    4.5.7 Moral Responsibility ......................................................................... 182
    4.5.8 Moral Perception ............................................................................... 182
    4.5.9 Action Guidance ................................................................................ 184
    4.5.10 Moral Motivation ............................................................................. 189
  4.6 Conclusion ................................................................................................... 191

CHAPTER 5:  INTENTION, COMMITMENT, AND COMMITTED INTENT .... 192
  5.1 Introduction .................................................................................................. 192
  5.2 Intentions and Commitments ...................................................................... 193
  5.3 Committed Intent ......................................................................................... 205
    5.3.1 Dent’s Moral Psychology of the Virtues .............................................. 206
    5.3.2 Sense-Desire ..................................................................................... 207
    5.3.3 Passional-Desire ............................................................................. 208
    5.3.4 Rational-Desire ................................................................................ 215
    5.3.5 Regulating Desire ............................................................................. 218
    5.3.5.1 Regulating Sense-Desire .............................................................. 219
    5.3.5.2 Regulating Passional-Desire ....................................................... 221
    5.3.6 Committed Intent ............................................................................... 221
  5.6 Conclusion ................................................................................................... 225

CHAPTER 6:  MEETING ANSCOMBE’S DEMAND ...................................... 226
  6.1 Introduction .................................................................................................. 226

BIBLIOGRAPHY ..................................................................................................... 234
ACKNOWLEDGMENTS

I wish to thank W. David Solomon, Alasdair MacIntyre, Michael DePaul, Alfred Freddoso, David K. O’Connor, Mary Keys, Fred Dallmayr, Edward Goerner, Lynn Joy, and all the faculty of the Department of Philosophy at the University of Notre Dame. In addition, I would like to acknowledge the organizers of the 2002 Oxford University Graduate Philosophy Conference and Sabina Lovibond for the valuable comments she offered on an early draft of Chapter 2. Also, my sincere thanks to my fellow graduate students, especially James Kruger, Chris Toner, Noel Birondo, Christian Miller, Becky Stengl, Kevin Elliot, Ben Huff, Kevin Connor, Chris Mirus, Eric DePlace, Alissa Branham, Sean Walsh and James Rissler. Finally, without the patience, encouragement, love and support of my wife, Amy, this dissertation might never have been completed.
1.1 Introduction

Being virtuous involves the possession of stable character traits that provide an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability reliably to act for, or because of, those reasons. This understanding of being virtuous presupposes that human beings can possess stable character traits with these capacities and abilities despite the serious and inescapable limitations of human practical reason. I call the truth of these presuppositions the psychological conditions of possibility for being virtuous. If one can provide an account of how it is possible\(^1\) for human beings to have stable character traits with these capacities and abilities, then one will have provided an aid to the virtue theorist.

However, providing for the psychological conditions of possibility for being virtuous will not suffice to vindicate virtue ethics because these conditions contain a presupposition of their own. In order for a character trait to provide an agent with a capacity consistently to recognize an objectively appropriate reason for action, an objectively appropriate reason for action must exist. Therefore, an account of how it

\[\text{It should go without saying that I am interested in how human beings can be virtuous in this world. I am not interested in the potential capacities and abilities of agents of other possible worlds. Nor am I interested in the potential capacities and abilities of agents of this actual world who are not human. I am interested in the potential capacities and abilities of human beings of this actual world.}\]
is possible for a human being to be virtuous presupposes the existence of objectively
appropriate reasons for action. I call the truth of this presupposition the normative
condition of possibility for being virtuous. If one can provide good arguments
concerning the source and nature of objectively appropriate reasons for action, then
one will have provided additional aid to the virtue theorist.

1.2 The Normative Condition

Virtue theorists provide a number of different accounts of the objectivity of
reasons for action by providing different descriptions of the relationship between, on
the one hand, virtues and the good, and on the other hand, virtues and right action. It
is appropriate at the outset to provide a brief overview of some of these accounts and
how they relate to one another. In addition, providing such an overview will present
an opportunity to introduce some technical distinctions that will prove useful in the
discussion that follows.

First, virtue theorists are often concerned with the relationship between the
virtues and the good, irrespective of any further specification of the good. Two views
are dominant: a teleological view and a non-teleological view. A teleological view
defends one of the following claims: 1) the virtues are non-constitutive means to
some independent good; 2) the virtues are constitutive means to and constitutive parts
of some good; or 3) the virtues are entirely constitutive of the good. The first view is
a consequentialist or utilitarian virtue theory which argues that the virtues are
valuable because they contribute to the maximization of preference satisfaction,
pleasure, or to any end that can be described as good independently of any direct
reference to the virtues. The second view is a neo-Aristotelian view which argues
that the good of ethical significance is a certain kind of life and that this life necessarily includes as an essential part the possession of the virtues. The third view identifies the possession of the virtues with the good and the good with the possession of the virtues. I will be calling a view of the third type a stoic view. One way to understand the differences between the three views is by saying that on the utilitarian view virtue is merely a means to happiness, that on the neo-Aristotelian view virtue is a necessary constitutive part of happiness, and that on the stoic view virtue is sufficient for happiness. According to any teleological view the good is taken to be ethically fundamental and virtue is treated as a derivative concept.

On a non-teleological view “the virtues and other internal properties of the agent [are] ethically fundamental, and the good is treated as a derivative concept.”

In a non-teleological theory, a motivation or other internal state is taken to be of fundamental value and any other conceptions of the good or value are derived from these prior sources of value.

There are two dominate ways in which virtue theorists relate virtue and right action: an agent-based view and a good-based view. According to an agent-based view “the evaluation of human actions is entirely derivative from and dependent on what we have to say ethically about [the inner life of] the agent who performs those actions.” Additionally, “the motivations or behavior of virtuous persons is what

---


makes an act right.” On a good-based view, the evaluation of human actions is entirely derived from and dependent on what we have to say about the good. However, virtue is regarded as either a mere means to right action or as a constitutive part of right action. In either case, though, the evaluation of the action is ultimately based on what is regarded as the good. As with the teleological view discussed above, there are at least three different things that could be said about the good.

Combining these various views we can construct the following table:

**TABLE 1.1: APPROACHES TO THE NORMATIVE CONDITION**

<table>
<thead>
<tr>
<th>Virtue and Right Action</th>
<th>Non-Teleological</th>
<th>Teleological</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agent-Based</em></td>
<td><em>Non-Teleological Agent-Basing</em></td>
<td><em>Util.</em></td>
</tr>
<tr>
<td><em>Good-Based</em></td>
<td><em>Util.</em></td>
<td><em>Utilitarian Agent-Basing</em></td>
</tr>
<tr>
<td><em>Neo-Arist.</em></td>
<td><em>Neo-Aristotelian</em></td>
<td></td>
</tr>
<tr>
<td><em>Stoic</em></td>
<td><em>Stoicism</em></td>
<td></td>
</tr>
</tbody>
</table>

In Chapter 2 I will be concerned primarily with the normative condition of possibility for being virtuous. I will be arguing that non-teleological agent-based virtue ethics is unable to provide for the normative condition of possibility for being virtuous. In addition I will be arguing that neo-Aristotelian agent-based virtue theory (sometimes called an agent-prior theory) is unhelpfully redundant and conceptually confused.

---

The reader will notice that there are a number of possible accounts of the normative condition of possibility. I will be dealing explicitly with only two and almost entirely in a negative manner. My objective is to provide an argument in favor of rejecting any form of agent-basing as an appropriate account of the normative condition of a virtue theory. I suspect that the arguments I supply against a neo-Aristotelian agent-based virtue theory could be redeployed (with appropriate amendments) against utilitarian agent-based virtue theory and that the argument I supply against non-teleological agent-based virtue ethics could be redeployed against a stoic agent-based virtue ethics.

As I said, for human beings, the possibility of being virtuous depends on three preconditions: one, the objectivity of appropriate reasons for action; two, the capacity for human beings consistently to recognize these objectively appropriate reasons for actions; and three, the ability for human beings to act reliably for, or because of, these objectively appropriate reasons. The two latter preconditions differ from the first in that the latter two are centrally claims about possibilities concerning human psychology. Even if one were to grant the objectivity of appropriate reasons for action, one could still object that being virtuous is impossible for human beings because they are not able consistently to identify these objectively appropriate reasons for action nor are they able reliably to act for, or because of, them.

1.3 The Psychological Condition

Richard Brandt has pointed out that “a necessary condition of an act’s being virtuous is that it ‘be based on fixed and permanent quality’ in an agent’s character. Obviously, this view makes no sense if the concept of a trait of character …makes no
One way to begin to clearly understand a concept is to inquire into what explanatory role those who use the concept want it to play. While many philosophers deploy the concept of a trait of character, it is of central interest to contemporary virtue theorists. We might ask, then, what role virtue theorists see the concept of a trait of character playing in an adequate virtue ethics. An inquiry directed to answer this question will generate a set of functional roles an account of which a theory of character must provide. These functional roles will provide the framework for an account of the psychological conditions of possibility for being virtuous.

1.3.1 The Functional Roles of Virtues

In 1958, Elizabeth Anscombe articulated an objection to both Kantianism and utilitarianism that laid the groundwork for her advocacy of virtue ethics. Her objection was that the moral law, a central concept of both Kantianism and consequentialism, is unintelligible without the notion of a lawgiver. Since modern moral philosophy has rejected the idea of a moral lawgiver, it cannot consistently appeal to the concept of the moral law. Therefore, we must reject and replace the dominant modern moral theories, Kantianism and consequentialism, with a theory whose central concepts are intelligible. She suggested as a viable alternative a theory in which the virtues are central.


Anscombe’s was only the first in a series of contemporary critics of Kantianism and consequentialism who turned to virtue ethics. Alasdair MacIntyre turned toward an Aristotelian inspired virtue ethics after noting the failure of what he calls the Enlightenment Project to provide a theoretical framework for the operation of rules in the absence of teleology. Bernard Williams at least suggests that a virtue theory is the best option “for us” because it is not an attempt to reduce the ethical to notions of obligation, duty, or outcomes. John McDowell, following a Wittgensteinian observation concerning the uncodifiability of rules, thinks that we must abandon Kantianism and consequentialism because the very idea that there are rules that could capture the requirements of moral behavior is unintelligible. Still others note that even if there were such rules, they cannot or do not have the practical implication they are purported to have.

These critics of modern moral philosophy see in the virtue tradition the resources for a better account of human practical reasoning. The central problem with modern moral theories is that they have failed to provide a realistic account of human practical reason. Williams argues that Kantian accounts of practical reason

---


require equating reflection with detachment from personal interests and desires. However, since human reflection and deliberation is essentially first personal, Kantian practical reason appears to be impossible for us.\(^12\) Foot argues that Kantian practical reasoning does not, in fact, issue in moral considerations that have automatic reason giving force. Without this feature, the Kantian account seems to collapse. Charles Taylor and many others have argued that consequentialism has reduced all human practical reasoning to mere instrumental reasoning.\(^13\) An account of human practical reason that reduces all practical reasoning to the application of rule, a decision procedure, or to mere instrumental reasoning is, according to the virtue theorists, inconsistent with central features of human rationality.

Contemporary advocates of virtue ethics believe that virtue ethics, in contrast to modern moral philosophy, provides a better account of practical reasoning that is consistent with the kind of limited rational beings we are. Since virtue theorists believe that traits of character play a central role in practical reason, we can conclude that one functional role a trait of character must play is to operate in human agency. In addition, as Brandt noted, a trait of character, as it operates in human agency, must be stable and entrenched.

Other virtue theorists provide a different but related critique of Kantianism and consequentialism. They think that both theories generate a morally unacceptable

\(^{12}\) Williams, *Ethics and the Limits of Philosophy*. See Christine M. Korsgaard, “Skepticism about Practical Reason” in *Creating the Kingdom of Ends* (New York: Cambridge UP, 1996), 311-334, for a reply to Williams’s argument in which she argues that Williams’s characterization is based on the assumption that there are no unconditioned principles of reason.

fissure between one’s motives and one’s judgments concerning right and wrong, which results in either moral disharmony if it is not corrected, or a morally impoverished life if it is corrected. Virtue ethics attempts to provide an account of morally harmonious virtuous activity according to which one is moved by what he judges to be right and judges to be right what moves him. Since morally harmonious virtuous activity is the result of an agent’s stable and entrenched traits of character, we can conclude that another functional role of traits of character is to produce moral harmony.

Virtue theorists also regard traits of character as playing a constitutive role in determining whether one achieves eudaimonia. Eudaimonia involves as a necessary, though not sufficient condition, an adequate amount of narrative unity in a human life. A way traits of character are thought to play a constitutive role in eudaimonia is by promoting the narrative unity of a human life.

In addition to asserting that traits of character operate in human agency, are stable and entrenched, provide for moral harmony, and are a constitutive part of eudaimonia, virtue theorists are often compelled to claim that an agent is morally responsible for the acquisition and modification of his or her traits of character. Therefore, an adequate virtue ethics must explain how it is possible for any agent to be held morally responsible for the state of his or he character.

Virtue theorists also look to traits of character to provide or facilitate moral perception, a reliable awareness of or sensitivity to the moral demands to which one

---

14 Stocker, “The Schizophrenia of Modern Ethical Theories.”

15 MacIntyre, After Virtue.
is subject in a given situation. The virtuous agent is thought to have the ability to see salient features of a moral situation that are not obvious to those who lack the virtues. Since virtue theorists attribute this moral perception to the possession of the virtues, the application of the concept of a trait of character must help explain how this kind of moral perception is possible.

Some critics of virtue ethics argue that the theory does not have the resources to guide or motivate action. The action guiding objection claims that without direct appeal to exception-less rules which are easily understood by moral agents, a moral theory is of no practical significance because it cannot tell us what we ought and what we ought not to do. Since it is assumed that virtue ethics does not have the ability to make an appeal to exception-less rules, it cannot provide moral guidance. The moral motivation objection typically takes two forms. The first version of the objection claims that virtue ethics cannot provide an account “of acting from (a sense of) duty, on or from (moral) principle, because you think you (morally) ought to, or are (morally) required to, or because you think it’s (morally) right.”

16 McDowell, “Virtues and Reasons.”


18 Hursthouse, On Virtue Ethics, 121.
of the moral motivation objection is that virtue ethics can give no account of why we should care about possessing the virtues, or about being a good human being.

Replies to these objections are necessary to begin to establish the adequacy of virtue ethics. The virtue theorist’s initial reply is to point out that it is not the theory that provides guidance or motivation. Rather, the virtues provide guidance and motivation. Unfortunately, virtue theorists have failed to provide a good account of how the virtues do this. Therefore, one requirement for an adequate virtue ethics is that virtue theorists provide a moral psychology of character that can account for moral guidance and motivation. In addition, recognizing this requirement provides an initial understanding of the concept of a trait of character; a trait of character must explain how virtuous action is possible.

The preceding description of the constraints on an adequate virtue ethics reveals a number of roles that the concept of a trait of character is intended to play not only in the virtue ethics as a theory, but more importantly in the moral psychology of human agents. The constraints therefore also generate an initial list of psychological functional roles that a trait of character must fulfill. These initial functional roles of traits of character are: 1) that traits of character operate in human agency; 2) that traits of character are relatively stable and entrenched; 3) that the stability of entrenched traits of character contribute to moral harmony; 4) that traits of character contribute to narrative unity; 5) that traits of character are constitutive of eudaimonia; 6) that traits of character are subject to voluntary acquisition and modification; 7) that the voluntary acquisition and modification of traits of character allow for the possibility that an agent is morally responsible for the status of his or her character; 8) that traits
of character promote moral perception; 9) that traits of character guide action; 10) that traits of character properly motivate an agent. Given these assumptions commonly made by virtue theorists concerning the functional roles of the virtues, I will be asking whether it is possible to construct an account of character within the theoretical constraints provided by these functional roles.\textsuperscript{19}

1.3.2 Anscombe’s Demand

While providing her critique of modern moral philosophy and suggesting virtue ethics as a better alternative, Elizabeth Anscombe also understood that the adequacy of virtue ethics depends on an adequate philosophical psychology. Therefore, given the weakness of Kantianism and consequentialism and the realization that an adequate philosophical psychology is a prerequisite to doing virtue ethics, Anscombe demands that we stop doing ethics until we have an adequate philosophy of mind, one that provides an understanding of intention, pleasure, desire, want, practical reason, and an analysis of virtue.\textsuperscript{20} Unfortunately, despite their commitment to the psychological roles presented above, few contemporary virtue ethicists have taken her advice.\textsuperscript{21}

\textsuperscript{19} It is not, however, part of my present project to ask whether these constraints are justified. Also, for the sake of keeping the argument manageable I have omitted any discussion of the role of traits of character is shaping and directing proper emotional or affective responses. However, I think that the account I develop in the dissertation can easily be expanded to incorporate an account of proper emotional responses.

\textsuperscript{20} Anscombe, “Modern Moral Philosophy.”

\textsuperscript{21} N. J. H. Dent makes explicit reference to this passage from Anscombe in his preface, stating that his book, \textit{The Moral Psychology of the Virtues} (Cambridge: Cambridge UP, 1984) is an attempt to provide this account of moral psychology. Hursthouse claims in \textit{On Virtue Ethics} that it appears to just be an historical accident that Dent’s study has not received more attention by ethicists.
1.3.3 Failures to Meet Anscombe’s Demand

Many ethicists and theorists deploy the notion of a virtue in their work without giving any substantial account of what a virtue is when understood as a psychological feature of human beings. Those that do try to give some indication of what a virtue is usually end up providing one of the following: some indication of what a virtue does but not how it does what it does; an examination of the areas of human moral and social life in which the virtues operate; an unhelpful declaration

---

I suspect that it is not just an historical accident, but rather the result of an unrecognized failure to take seriously the moral psychology on which a virtue theory depends.

22 See Georg Henrik von Wright, Varieties of Goodness (Routledge, 1963; repr., Bristol: Thoemmes Press, 1996). Because von Wright has the view that good and bad acts are secondary to good and bad intentions that are in turn secondary to notions of harm and benefit, his account of virtue emphasizes not the psychology of the virtues, but their utility. That is, von Wright’s account of the virtues concentrates on the role they play (whatever they are) in counteracting the ‘obscuring’ effects of passion upon an agent’s judgment of harms and benefits. See also Philippa Foot’s account of the virtues in “Virtues and Vices,” in Virtues and Vices and Other Essays in Moral Philosophy (Berkeley: University of California Press, 1981). Here she is concerned to show that virtues are corrective, but of two kinds. Some virtues operate by helping an agent to resist temptation while other virtues operate to bolster weak motivation. While both von Wright and Foot refer to the (assumed) relationship between virtues and the will, choice, reason, and knowledge, they do not provide any account of a virtue as a psychological feature of a human being. In addition, McDowell in “Virtue and Reason” defines virtue as an ability to recognize requirements that situations impose on behavior. See finally Alasdair MacIntyre, Dependent Rational Animals, (Peru: Open Court Publishing Company, 1999) where he describes virtues as qualities of mind and character that allow someone to recognize and acquire certain human goods. Here we see the motivation for certain criticisms of MacIntyre’s work that charge that his account of the virtues is utilitarian in some way. For an example of this criticism see Gregory Velazco y Trianosky, “What is Virtue Ethics All About?” in Virtue Ethics: A Critical Reader, ed. Daniel Statman (Washington, D.C.: Georgetown UP, 1997), 42-55. Of course, on MacIntyre’s view the virtues are not merely instrumental, but are constitutive of a certain way of life, and thereby constitutive of a good life.

that a virtue is a character trait;\textsuperscript{24} a declaration that the virtues are related to \textit{eudaimonia};\textsuperscript{25} or a hint of a psychological account of the virtues that does not tell us enough to allow us to construct a realistic moral psychology to serve as a foundation of virtue ethics.\textsuperscript{26}

While these methods of discussing the virtues are helpful as far as they go, they each fail to meet Anscombe’s Demand. Each fails to provide a realistic psychological account of the virtues. Yet, having provided little that is helpful with regard to determining the psychology of the virtues, these contemporary authors jump head long into discussions of virtuous action, the merits of virtue theory to other rival alternative theories, and into defenses of their position against the objections raised by the advocates of rival theories. One is left wondering how it is possible to provide an account of virtuous action, let alone champion the cause of the revival of virtue ethics over other theories without first giving a psychologically realistic account of what a virtue is.

In suggesting that what a virtue theory needs is a psychologically realistic account of what a virtue is, I want to insist that while recapturing the meaning of ‘virtue’ is an important step in the development of a virtue theory, it is not sufficient


\textsuperscript{25} This may be the same as the first method, but I single it out here because Hursthouse, in “Virtue Theory and Abortion,” explicitly states that the virtues are defined in terms of eudaimonia.

for a cogent and defensible position. As Hursthouse notes, the philosophical community has not reached a familiarity and comfort with the meaning of ‘virtue’ as they have with the concepts of other theories. Whatever the status of the meaning of ‘virtue’ within the philosophical community, virtue theorists must also make progress in articulating to what they think the concept ‘virtue’ refers. A theory which claims that virtue terms are more basic than deontic terms, but does not attempt to indicate what the reference of the virtue term is, is not really a virtue theory at all, at least in the sense in which I shall use the term. This is because a virtue theory presupposes a rich moral psychology, and if virtue theory is to survive as a real option, it must articulate and defend this moral psychology. We must meet Anscombe’s Demand.

It is perhaps misleading for me to suggest that contemporary virtue ethicists have not provided any indication of what kind of psychological feature a virtue is. However, attempts by contemporary virtue theorists are, I think, not entirely helpful and perhaps even misleading in important respects. The tradition of the virtues has often described a virtue as a ‘habit.’ However, in contemporary English, ‘habit’ has the connotation of referring only to actions that are performed without thought.

---

27 I take it that one of the central tasks of MacIntyre’s *After Virtue* is to provide a discussion of the changing meanings of virtue through history and develop a re-tooled meaning that is perhaps applicable to the contemporary moral predicament.

28 See Hursthouse, *On Virtue Ethics*. But see also her “Virtue Theory and Abortion” in which she explicitly accuses the concepts of deontology and utilitarianism of being equal in obscurity to the concepts deployed in a virtue theory.


30 See von Wright, *Varieties of Goodness*.
will the common use of ‘disposition’ help matters because it indicates a tendency to perform certain actions.\footnote{See von Wright, \textit{Varieties of Goodness}. See also Foot, “Virtue and Vice” in \textit{Virtues and Vices} where she describes a virtue both as a disposition and as a characteristic of a person, and Jorge Garcia, “The Primacy of the Virtuous,” \textit{Philosophia} 20, no. 1-2 (1990): 69-91 where he defends the virtue theorists against the efforts of Kantians and utilitarians to subordinate the virtues by reducing them to conscientiousness or benevolence (i.e., dispositions to do what is right).} ‘A state of character’ is perhaps more helpful in that it implies that virtue refers in an important sense to the psychological make-up or composition of an individual. To refer to a state of character when deploying the concept of ‘virtue’ indicates that what we mean by virtue is something that makes a person the kind of person he or she is.\footnote{See Zagzebski, \textit{Virtues of the Mind}; MacIntyre, \textit{After Virtue}; and Julia Annas, “Self-love in Aristotle,” \textit{Southern Journal of Philosophy} 27, supp (1988): 1-18.} But ‘state of character’ itself needs a psychological explanation.

1.3.4 A Sketch of a Strategy to Meet Anscombe’s Demand

My task in this dissertation is to provide a psychological explanation of the possibility of being virtuous, an explanation which entails a prior psychological explanation of character. By character I mean the mental architecture which makes the possession of the virtues possible. I will argue that having character, as I will define it, is a necessary, but not a sufficient condition for being virtuous. If an agent possesses a stable, hierarchical deliberative structure with the following capacities and abilities, then she can be said to possess character: 1) to desire and to have beliefs; 2) to consider her desires in terms of past experience and expectations of future consequences; 3) to choose between various possible courses of action by weighing the possibilities in terms of their expected future consequences; 4) to
coordinate her actions over time via the formation of acceptances, intentions, plans, and policies; 5) to recognize herself as a temporally extended creature and anticipate future regret; 6) to arrive at higher order attitudes concerning the priority to be given to the range of plan-constrained lower order reasons for action; 7) to take a stand with respect to the decisive importance of particular reasons for action via the formation of self-governing policies; and finally, 8) to have first-order reasons for action reflecting the commitments incorporated in her self-governing policies. An agent with these capacities has character. If an agent with character is committed to what is good and best, then this agent is virtuous.

While I will have much more to say about the distinction between character and virtue, let me reiterate my earlier claim that having character (as understood in this dissertation) is a necessary, not a sufficient condition for being virtuous. An agent who has character can fail to be virtuous. The role of character in my account is to distinguish those who meet the psychological conditions of possibility for being virtuous from those who are morally weak, easily tempted, or easily manipulated. An agent who has character is an agent with integrity, an agent whose mental architecture is stable and strong, and an agent whose commitments to objects she regards as of decisive importance form, shape and focus her deliberative framework. Only an agent who has character is a candidate for being judged to be virtuous. Whatever else we might say about the continent, the incontinent, the beastly, the wanton, the impulsive, and the impatient, we would never say that they are candidates for being judged virtuous.

I will suggest that having character is not, properly speaking, a mental state; it is the result of a set of complex, stable, and temporally extended relationships among a host of mental states, including beliefs, acceptances, desires, intentions, plans and policies. The task of a moral psychology of character is to identify and describe these relationships. I will argue that Michael Bratman’s temporally extended planning theory of human agency provides the resources for a promising description of the relationships involved in having character. I follow Bratman in arguing that an agent’s commitments can become embedded in her intentions and plans in such a way that her commitments take on a central role in directing her attention, framing her practical reasoning, and in motivating her to pursue certain ends. Due to the profound limitations of human rationality, the deliberative framework created by an agent’s intentions and plans exerts pressure on her to maintain consistency and coherence within her practical deliberations. I argue that the deliberative framework characteristic of a temporally extended planning agent, and the resulting consistency and coherence of practical deliberation, provide a fruitful description of the relationships between the mental states constituting character. When this account of the moral psychology of character is coupled with an adequate account of the normative condition of possibility for being virtuous (in the form of a normative

---

34 Intentions play a central role in other discussions of virtue and character. For example see Robert Audi, *Moral Knowledge and Ethical Character* (New York: Oxford UP, 1997), and Garcia, “The Primacy of the Virtuous.” Garcia’s account is importantly different from the one discussed here because the role of intention is to provide a means of providing an independent account of virtuous action. That is, on his account certain intentions are virtuous. I do not intend to discuss which intentions are virtuous or vicious, though I agree that certain intentions are virtuous. What I will be discussing is the role future-directed intentions play in the possession of character.

35 I share this view of the task of moral psychology with Dent. See Dent, *The Moral Psychology of Virtue*. 
theory concerning the good), virtue theory stands a chance of becoming a genuine rival moral theory.

1.4 A Wittgensteinian Objection

The account of character that I present in the following pages I now recognize to be open to a number of possible objections that stem from the later philosophy of Ludwig Wittgenstein. It might have been prudent to avoid a confrontation with the Wittgensteinian in these pages. However, there is one significant reason why I cannot avoid such a confrontation. The burgeoning tradition of modern virtue ethics is heavily influenced by philosophers to whom Wittgenstein provides an important touchstone concerning issues of both philosophical psychology and the philosophy of language. Of the major contemporary advocates of neo-Aristotelian virtue ethics, nearly all credit Wittgenstein as having made important contributions to their views. Consider, for example, Elizabeth Anscombe, Philippa Foot, Peter Geach, Rosalind Hursthouse, Alasdair MacIntyre, Bernard Williams, John McDowell, Anthony Kenny, Sabina Lovibond and Georg Henrik von Wright, just to name a few. There is an interesting history yet to be written that tells the story of Wittgensteinian reactions to the debates concerning meta-ethics during the 1950’s and 1960’s and the subsequent revival of virtue ethics. However that historical story is told, the fact is that many virtue ethicists will be resistant to my project because they are Wittgensteinians concerning matters of philosophical psychology. My initial reaction to why this is so is that it is a result of contingent historical fact. There do not appear to be any necessary connections between Wittgenstein’s philosophy of psychology and virtue ethics, nor vice-versa. Nonetheless, I think it is important to provide the
Wittgenstein objection to the view of philosophical psychology I provide in this
dissertation.

I also believe that it is important to provide an initial reply to this kind of
objection. From the perspective of my Wittgensteinian allies, this reply will be
inadequate in the extreme, but it will at least place my project in proper relief and
help make it clear how my project can at least fill a gap in virtue ethics given a set of
assumptions shared by most enemies of virtue ethics, though perhaps not by many of
its adherents. I do not believe that those of us who find virtue ethics attractive want at
this stage of the debate to be put in the position of requiring one to become a
Wittgensteinian concerning psychology before he or she can understand and perhaps
come to accept virtue ethics as a reasonable moral theory.

Unfortunately, I have placed myself in the middle of (at least) a two front
battle. Behind me I have Wittgensteinian virtue ethicists who will resist the
suggestion that an account of the moral psychology of character requires an appeal to
dispositions understood as actual psychological features of an agent, particularly if
these psychological features are understood as playing a causal role in an agent being
virtuous. On the other hand, my intention is to provide an account of character that
can meet the kind of objections being presented to virtue ethicists by those who want
an account of the moral psychology of the virtuous agent provided in terms that they
can understand, terms which usually involve desire-belief pairs and mental causes.
But since the Wittgensteinian rejects the terms in which the critic wants the account
to be provided, the Wittgensteinian virtue ethicist is tempted to make the kind of
claim that Rosalind Hursthouse makes when recounting her frustration when
attempting to discuss the moral psychology of virtue without appeal to drives, beliefs, desires, and mental causes: “But you don’t believe that so-and-so if you’re a Wittgensteinian.”

Hursthouse hopes that making this Wittgensteinian reply early on in a defense of virtue ethics will help clarify where the disagreement lies between those sympathetic to Wittgenstein’s philosophical psychology and the majority of analytic action theorists who believe that “philosophy is supposed to uncover or construct the foundations of our thought.”

I agree with Hursthouse that it is an important intellectual activity to clarify where the disagreement lies between the Wittgensteinian and the analytic philosopher concerning the nature of the moral psychology of character. To this end I will on occasion be making explicit where I see the disagreement lying. In addition, I will sometimes attempt to show where there is reason to believe that the disagreement can be mitigated. In order to bring to light the Wittgensteinian objections to my project, I will be providing a brief discussion of Wittgenstein’s position concerning the status of mental states.

36 I think the entire passage is interesting and worth quoting at length: “In contemporary philosophy of action, there is a fervid debate about whether any intentional action must be prompted in part by a desire, or whether it is possible to be moved to action by a belief…. The debate all takes place against the background of the assumption that belief and desire are as different as gold and oxygen, and usually also the assumption that the distinction between the rational and the non-rational is equally hard and fast.…

“To anyone sympathetic to the writings of the later Wittgenstein… rejections of clear-cut distinctions in philosophical psychology are as natural and necessary as breathing…. I have found that when teaching virtue ethics to graduate students, or discussing papers… with fellow philosophers,… what often blocks understanding is the unconscious assumption [on my audience’s part] that everyone shares the view that, for example, beliefs and desires are natural kinds, or that a reason [for action] is a belief/desire pair that causes an action, or that all mental states are brain states – or, more generally, that philosophy is supposed to uncover or construct the foundations of our thought. Struggling to square these assumptions with what is said, the audience finds what is said deliberately obscure or willfully incomplete, or inconsistent, or open to such blindingly obvious objections that they think they cannot have understood. Sometime – not always, of course – the cloud lifts if one says, ‘But you don’t believe that so-and-so if you’re a Wittgensteinian’” (Hursthouse, On Virtue Ethics, 16).

37 Hursthouse, On Virtue Ethics, 16.
According to Wittgenstein, we are often deceived by our uses of psychological concepts.\textsuperscript{38} For our purposes here, we are deceived when our uses of psychological concepts appear to make reference to mental states and mental apparatuses. In general, Wittgenstein believes that using psychological concepts in ways that appear to refer to mental states and mental apparatuses generates philosophical confusions concerning the nature of the mind, particularly when we try to provide a complete and accurate description of how we use psychological concepts. Despite our mastery of the uses of psychological concepts in our own language, our ignorance concerning the proper description of these uses tempts us to appeal to philosophical fictions, such as mental states and mental apparatuses, on those rare occasions when we actually attempt to describe our use of these psychological concepts.

Malcolm Budd has emphasized two specific reasons why Wittgenstein was particularly wary of our attempts to describe our use of psychological concepts.\textsuperscript{39} First, in our descriptions of our uses of psychological concepts we often are strongly tempted to attempt to explain what we think is happening in the mind that justifies our using our psychological concepts in the ways we use them. The danger lies in the fact that by attempting to justify our using these concepts as we do by explaining what we believe is happening in the mind, we are inclined to invent objects, events, and processes to which our words refer. The second reason Budd mentions is that psychological concepts have particularly ramified uses; that is, the use of a given

\textsuperscript{38} Malcolm Budd, \textit{Wittgenstein’s Philosophy of Psychology} (New York: Routledge, 1989).

\textsuperscript{39} Budd, \textit{Wittgenstein’s Philosophy of Psychology}, 6.
psychological concept has so many related uses such that it is an incredibly difficult task to give an accurate description of its uses. Unfortunately, rather than recognizing the difficulty of the task and being diligent in our investigations into the actual uses of a psychological concept in our language, we instead present a simplified account of its uses that is under constant pressure to account for all the apparent exceptions to the account.

Not only does the ramified use of a particular concept contribute to the difficulty of providing a sufficient description of the uses of psychological concepts, but the ramified uses of different kinds of psychological concepts also contribute to this difficulty. Psychological concepts can be grouped in one way due to a certain similarity in use while not grouped together based on another similarity in use. For example, a distinction Wittgenstein makes which is relevant to this dissertation is the distinction between, on the one hand, states of consciousness, and on the other hand, dispositions (or what he sometimes calls capacities, mental states or dispositional psychological states). Budd argues that according to Wittgenstein the distinction between the two kinds of psychological concepts is ultimately based on the fact that it makes sense to say that states of consciousness have a genuine duration nor to say that they “have endured continuously from one time to another.” If it does not make sense to speak of a psychological condition as enduring for a definitive time, then this restriction on the use of the corresponding concept designates that psychological condition as a disposition. According to this distinction, sensations, sense impressions and images are states of consciousness while knowledge, belief,

40 Budd, *Wittgenstein’s Philosophy of Psychology*, 14.
understanding, and intention are dispositions. Therefore, according to Wittgenstein, dispositions, capacities and mental states are those psychological concepts which it does not make sense to speak of as enduring for a distinct or genuine duration. However, despite the similarity of duration between states of consciousness, there are multiple ways in which concepts that are designated as states of consciousness can be further distinguished based on further dissimilarities in use. Therefore, each kind of concept is further ramified despite certain kinds of similarities. A proper description of the uses of a psychological concept would require that one account for all these levels of ramified use.

Wittgenstein’s fundamental position is that the task of philosophy is to carefully and thoroughly describe the actual uses of words in language in order to discover the meaning of these words. As Wittgenstein says at Philosophical Investigation 109, “Philosophy is the battle against the bewitchment of our intelligence by means of language.” In Wittgenstein’s view, language is both the source of our bewitchment and the means by which we can escape. Philosophical problems are the result of being tempted by the features of our language to justify our use with an explanation of that to which our language appears to refer. This mistake can only be remedied by a thorough investigation into the actual uses and functions that words play in our language. With a thorough understanding of the uses and functions of our words and concepts, philosophical problems will evaporate. Therefore, the task of the philosopher is merely to describe the uses to which we put our concepts, particularly psychological concepts.
An essential component of this task is a way of classifying similar and dissimilar uses of psychological concepts on the basis of what it makes sense to say in our language. One of these fundamental dissimilarities in use leads to Wittgenstein’s distinction between states of consciousness and dispositions which was introduced above. Recall that Wittgenstein identified knowledge, belief, understanding, and intention as dispositions. Wittgenstein’s objection to regarding intentions as states of consciousness begins with his claim that a state of consciousness has a genuine duration. An intention is not a state of consciousness because it is not possible for everything that the agent intends to be present to the agent’s mind at the instant that a predicate concerning his or her intentions become true of her. Wittgenstein’s objection is that the complete meaning that an agent intends at time t1 is not determined by what is present to the agent’s mind at t1 – and what A can truthfully say at t2 about his intentions at t1 is also not determined by what was before A’s mind at t1. The psychological condition of a person to whom a particular predicate concerning his or her intention is truly applied need not be distinguished from the condition of someone to whom a particular predicate is not truly made because the predication is not made by reference to what was present to the agent’s mind at a given, distinct time. So, Wittgenstein objects to the claim that intention is a state of consciousness because all that is entailed by intending cannot be present to the mind of the agent at the moment that the predicate becomes true of him or her.

Clearly, the same kind of argument could be presented against the claim that virtue is a state of consciousness. In order to make the claim that to be virtuous is a state of consciousness, one would have to defend the claim that all that being virtuous
entailed must be present to the mind of the virtuous agent anytime the predicate ‘is virtuous’ is true of him or her. However, for precisely the same reasons that everything one intends cannot be before one’s mind at any given time, so too all that is entailed by being virtuous cannot be before an agent’s mind at a given time.

However, we are still left with the option of considering intention (and virtue) as a disposition, as something it does not make sense to speak of as having a genuine duration. At *PI* 149, Wittgenstein appears to argue that it is inappropriate to claim that knowledge of the alphabet is “a state of a mental apparatus.” According to Budd, a mental apparatus in Wittgenstein’s usage is a disposition which is the *causal* foundation of a capacity. Wittgenstein also appears to assume that a mental apparatus is a *physical* state which operates in a *law-governed* manner to produce the manifestation of the capacity. At *PI* 149, Wittgenstein says,

> If one says that knowing the ABC is a state of the mind, one is thinking of a state of a mental apparatus (perhaps of the brain) by means of which we explain the manifestations of the knowledge. Such a state is called a disposition. But there are objections to speaking of a state of mind here, in as much as there ought to be two different criteria for such a state: a knowledge of the construction of the apparatus, quite apart from what it does.

Budd replies to this objection to considering knowledge (or any disposition, including an intention or virtue) as a state of a mental apparatus by granting that it is not necessary for one to observe the state of one’s own mental apparatus in order to discover that one has a particular type of knowledge. Budd also grants that knowledge of a state of a mental apparatus in the absence of knowledge of the
functional roles it allows me to fulfill is not sufficient to have knowledge of one’s capacities. However, Budd also says the following:

[It] is not the fact that my mental apparatus satisfies a certain physical description, but the fact that it is this state of my mental apparatus that enables me to reproduce the alphabet without aid that entitles us to say that my knowledge of the alphabet is founded in this state of my mental apparatus. It is also true that the capacity could be differently based in different people, or in the same person at different times: possession of the same capacity does not require a mental apparatus constructed in an identical fashion. But this does not imply that it is coherent to suggest that the capacity might be possessed by someone who lacks a mental apparatus in which the capacity is based. Wittgenstein appears to be relying on the supposition that if there is a single criterion for the possession of a psychological capacity [i.e., the ability to reproduce the alphabet], and this criterion makes no reference to a mental apparatus in which the capacity is founded, then it cannot be an a priori requirement that possession of the capacity must be grounded in a mental apparatus.41

Budd’s objection to Wittgenstein appears to be a functionalist objection. It is a weak thesis. It is a thesis about the possibility of the capacity. The possibility of the capacity requires the postulation of a mental apparatus in which the capacity is based. However, Budd need not say much, or anything, about the mental apparatus. However, he does appear to be committed to some causal relationship – though not necessarily one that can be the basis of knowledge of causal laws. Nonetheless, Budd’s objection does open the door for at least the initial move in my argument. It follows from Budd’s objection that there is nothing in Wittgenstein’s arguments that prevents one from identifying a number of functional roles, or capacities, which are

41 Budd, *Wittgenstein’s Philosophy of Psychology*, 27.
required to speak of someone as having the virtues, and then postulating that each individual must have a mental apparatus in which these capacities are grounded. Furthermore, if a discussion of the mental apparatus is not a discussion of physical states, but of functional states, and furthermore, the discussion can account for the possibility that different states of mental apparatuses can play the same functional roles, then many of Wittgenstein’s objections are mitigated if not dissolved. Therefore, it appears that if one is a functionalist about the nature of the mind, then one can accept Wittgenstein’s critiques of dualism, behaviorism, and reductive materialism while accepting a functionalist account of mind that will allow one to continue to speak intelligibly of states of mental apparatuses (mental states). This is precisely the strategy I will be pursuing in this dissertation.

1.5 Conclusion

I have started from the assumption that being virtuous involves the possession of character traits that provide an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability reliably to act for, or because of, those reasons. This assumption presupposes one normative condition of possibility and two psychological conditions of possibility for being virtuous. Following Anscombe, I submit that virtue ethics cannot be a cogent theory without a psychological account of character. In this chapter, I have identified from the working assumptions of leading virtue theorists ten initial functional roles that virtue theorists appear committed to accounting for within their moral psychology of character. An account of these functional roles will provide a basis for the psychological conditions of possibility for being virtuous. Despite possible resistance
from within the ranks of virtue theorists, my intended method is to use the concept of a future-directed intention and the temporally extended planning theory of human agency as a philosophical stepping-stone to a psychological understanding of ‘character.’ I will suggest that ‘a state of character’ is not, properly speaking, a mental state; it is the result of a set of complex, stable, and temporarily extended relationships among a host of mental states.

Before moving on to this task, I want to address in Chapter 2 a series of related questions regarding virtue ethics and the source of the objectivity of appropriate reasons for action. To give an account of how it is possible for objective reasons for action to exist is a first step in meeting Anscombe’s Demand.
CHAPTER 2

THE NORMATIVE CONDITION OF POSSIBILITY FOR BEING VIRTUOUS

2.1 Introduction

Being virtuous involves the possession of character traits that provide an agent with the capacity to consistently recognize objectively appropriate reasons for action as well as the ability to reliably act for, or because of, those reasons. In order for any particular character trait to provide an agent with a capacity to consistently recognize objectively appropriate reasons for action of a certain kind, objectively appropriate reasons for action of an appropriate kind must exist. Therefore, an account of how it is possible for a human being to be virtuous presupposes the existence of objectively appropriate reasons for action. I shall be calling the truth of this presupposition the normative condition of possibility for being virtuous. Therefore, a virtue theorist must provide a coherent account of how it is possible for objectively appropriate reasons for action to exist.

There are two dominate ways in which virtue theorists provide an account of objectively appropriate reasons for action, an agent-based account and a good-based account. According to the agent-based view “the evaluation of human actions is entirely derivative from and dependent on what we have to say ethically about [the
inner life of] the agent who performs those actions.”⁴² According to agent-basing, “the motivations or behavior of virtuous persons is what makes an act right.”⁴³ Therefore, the primary source of objectively appropriate reasons for action is objectively good inner states of the agent. On the good-based view, the evaluation of human actions is entirely derived from and dependent on what we have to say about the good. Therefore, the primary source of objectively appropriate reasons for action may have sources other than the objectively good inner traits of an agent. In this Chapter I will be concerned primarily with the normative condition of possibility for being virtuous. I will be arguing that non-teleological agent-based virtue ethics is unable to provide for the normative condition of possibility for being virtuous. In addition I will be arguing that with respect to the normative condition of possibility, teleological (neo-Aristotelian) agent-based virtue ethics is unhelpfully redundant and conceptually confused.

An effort to provide an account of the normative condition of possibility for being virtuous is in part a quest for an account of the minimal amount of knowledge required to allow for the possibility of human beings having the capacity to recognize appropriate reasons for action despite all their limitations. Whatever else this knowledge might be, it must involve some minimal conception of what is good or best. Some minimal conception of the good or the best is required in order to begin to speak coherently of appropriate reasons for action. Without a coherent way to talk about appropriate reasons for action, there is no coherent way to talk about virtue, ⁴⁴


being virtuous or right action. Nor is there any way to coherently speak about the psychological conditions of possibility for being virtuous. For one might argue that if such objectively appropriate reasons for action did exist, human beings would be able to consistently identify them and reliably act for, or because of, them, and yet deny that such objectively appropriate reasons for action do in fact exist.

My primary concern in Chapters Three and Four will be with the psychological conditions of possibility for being virtuous, accounts of which are necessary conditions for a plausible virtue ethics. In this Chapter I am concerned primarily with the normative condition of possibility. In particular I am eager to reject any agent-based account of the sources of reasons for action, especially insofar as this account provides a foundation for a theory of right action.

Agent-based virtue theorists typically define right action in terms of what the virtuous agent would do in the circumstances. I will argue in what follows that non-teleological agent-based virtue theorists make the same mistakes Euthyphro made when he attempted to define piety in terms of what is loved by all the gods:⁴⁴ he or she defends a definition of right action that provides no reason for action, believes that ‘being right’ is a counterpart property,⁴⁵ and has confused a non-essential feature of right action with the essence of right action.

---


⁴⁵ This term is borrowed from R. E. Allen, Plato’s Euthyphro and the Earlier Theory of Forms (London: Routledge and Kegan Paul, 1970). A counterpart property is a property that an action has only in virtue of the fact that an agent performs it. “Counterpart properties exist because of the activities, but the activities do not exist because of the counterpart properties” (Allen 40). The account of the Euthyphro presented in this chapter is heavily influenced by Allen’s account.
Furthermore, I will argue that more sophisticated versions of agent-based virtue ethics that attempt to define right action in terms of what the virtuous agent would do, and then define the virtuous agent in terms of something further, are only attractive when one fails to distinguish virtue, understood as a normative concept, from virtue, understood as a psychological concept. A failure to take this distinction seriously causes teleological agent-based accounts of right action to be conceptually confused and unhelpfully redundant.

Having discussed, in outline at least, the normative condition of possibility, we will then be in a better position to discuss the psychological conditions of possibility in an effort to move toward a moral psychology of character.

2.2 Reasons for Action

When Socrates asks Euthyphro, “What is piety?” our initial explanation of Socrates’ question is that Socrates wants to persuade Euthyphro that Euthyphro does not know what he claims to know concerning the nature of piety. In the context of an explanation of the purpose and form of early Platonic dialogues, our initial explanation is a good beginning. Socrates does indeed proceed to engage Euthyphro in a process of *elenchus* designed to demonstrate to Euthyphro that he does not possess knowledge concerning piety, despite his protestations to the contrary. This interpretation is familiar to us all. By engaging Euthyphro in a dialogue concerning the discovery of a definition of piety, Socrates is engaged in an epistemological exercise, namely, to find appropriate definitions for moral terms.

However, I wish to suggest a different description of Plato’s early dialogues, one that reveals that Plato was centrally concerned with the concept of a reason for
action. There is much to learn from Plato concerning the moral psychology of virtue and vice, but it requires a different approach to his discussions, one that de-emphasizes the role of the theory of forms and takes as central the question of formal criteria for acceptable reasons for action and the formal criteria for the possession of virtue and vice. In order to develop part of this view, I want to return to the *Euthyphro*.

In the *Euthyphro*, Socrates seeks an answer to the question, “What is piety?” Socrates is surprised that anyone would prosecute his own father based on ambiguous evidence such as Euthyphro has against his father. Only someone who is confident in his knowledge of piety would believe he had adequate reason to risk such a potentially impious action as prosecuting one’s own father. Euthyphro is a young man, an apparent expert concerning religious and legal matters, and a self-described seer, who claims to be accurate in his predictions, but whose predictions concerning Socrates’ future are false. He is about to indict his own father for murder based on the beliefs (a) that his father actually committed an unjustified homicide, and (b) that to fail to prosecute his father would be to allow ‘pollution’ to infect himself and his father. Euthyphro feels confident in his expert knowledge regarding what religion demands and what religion forbids, and so he believes he has unassailable knowledge concerning the question of whether his father committed murder. Euthyphro’s actions necessarily raise the problem of knowledge.

And thus the questions of the dialogue are introduced: what is piety and what standard is to be used to determine which actions are genuinely pious and which are not? On Euthyphro’s behalf, Socrates proposes that “What *all* the gods hate is
impious, and what they *all* love is pious.”46 After Euthyphro accepts this definition as an accurate representation of his position, Socrates asks him, “Is the pious loved by the gods because it is pious? Or is it pious because it is loved by the gods?”47 The discussion that follows appears to be either a discussion of active and passive voice or a discussion of the fact that activities require objects.

A better interpretation is that the discussion is one concerning conditional priority.48 Loving is an activity of an agent; being loved is a counterpart property of the object. A counterpart property is a property that an action has only in virtue of the fact that an agent performs it. Socrates’ claim is that the counterpart property exists because of the activity of the agent. The activity does not exist because of the counterpart property of the object. Action is conditionally prior to counterpart properties produced by action. Socrates offers a series of often baffling examples of the relationship between activities ($\alpha$) and counterpart properties ($\pi$) resulting in the general principles (M) and (N).

The Examples (10b):

E: $\neg$($X$ is seen ($\alpha$) because $X$ has the property ‘being seen’ ($\pi$))  
F: ($X$ has the property ‘being seen’ ($\pi$) because someone sees $X(\alpha)$)

G: $\neg$(Something leads $X$ ($\alpha$) because $X$ has the property ‘being led’ ($\pi$))  
H: ($X$ has the property ‘being lead’ ($\pi$) because someone leads $X$ ($\alpha$))

46 Plato *Euthyphro* 9d.  
47 Plato *Euthyphro* 10a.  
I: ~(Something carries X (α) because it has the property ‘being carried’ (π))
J: (X has the property ‘being carried’ (π) because someone carries X (α))

K: ~(Something becomes X (α) because X has the property ‘coming to be’ (π))
L: (X has the property ‘coming to be’ (π) because something creates X (α))

The General Principles (10c):

M: ~(Something affects X (α) because X has the property ‘being affected’ (π))
N: (X has the property ‘being affected’ (π) because something affects X (α))

Understanding the notion of the conditional priority of activity to properties,
Euthyphro asserts that the pious is loved by the gods because it is pious, not pious because it is loved by the gods. Why? Because ‘being loved’ is a counterpart property of the gods’ love for those things that are pious. The conditionally prior activity that produces the counterpart property, being loved, is the activity of loving.

We can understand Socrates’ application of the general principle to the case of loving in the following way:

Application of the General Principles to the case of piety (10c – d):

O: ~(X is loved (α) because X has the property ‘being loved’ (π))
P: (X has the property ‘being loved’ (π) because someone loves X (α))

Definition: let X equal piety
Definition: let the relevant property be ‘being loved by all the gods’
Q: \( \neg (\text{Piety is loved by all the gods } (\alpha) \text{ because piety has the property } \text{‘being loved by all the gods’ } (\pi)) \)

R: (Piety has the property ‘being loved by all the gods’ (\(\pi\)) because all the gods love it (\(\alpha\)))

The focus of this discussion is to demonstrate that the actions of an agent are always conditionally prior to the counterpart property of a patient. Therefore, a counterpart property can never be given as an explanation of or justification for the actions of an agent. Suppose you asked me why I love my wife. Moreover, suppose that I replied to your question by saying, “I love my wife because she has the property ‘being loved by Liam.’” Such a reply does not explain why I love her.

The explicit force of Socrates’ argument concerning counterpart properties is entirely negative. However, the positive implication of the argument is that an explanation of action must provide a reason for action. Consider again the case of my loving my wife. In order to explain my loving her I have to give my reasons for loving her, or more generally, my reasons for action.

Suppose you asked me why I came to South Bend today. And suppose further that I replied by saying, “I came to South Bend today because South Bend has the fortunate property ‘Liam here on Mondays, Wednesdays, and Fridays.’” That would not explain why I came today. What explains my going to South Bend are my reasons for coming, namely, to do some work, to fulfill my teaching responsibilities, to return a recalled book to the library, etc. Likewise, what explains my loving my wife are some further properties she possesses, such as intelligence, beauty, kindness, humor, faith, etc. The reason for my loving her is not that she is loved by me.
At 10d Euthyphro recognizes reasons for action among the gods when he admits that the gods love pious things because those things are pious.\textsuperscript{49} That something is pious is a reason to love that thing, just as that a thing is sweet is a reason to eat it, or that a thing is pleasurable is a reason to pursue it, or that a thing is painful is a reason to avoid it (these may not be good reasons, but they are nonetheless reasons for action). By 10d, Socrates has gotten Euthyphro to recognize that all explanations and justifications of actions require reasons and that counterpart properties do not count as reasons. However, the argument then takes an interesting turn. Euthyphro assents to the following premises (10d):

\begin{align*}
A: & \ (\text{The pious is loved by all the gods (}\alpha\text{) because it is pious (}\pi\text{)}) \\
B: & \sim(\text{The pious is pious (}\pi\text{) because it is loved by all the gods (}\alpha\text{)})
\end{align*}

But why does Euthyphro assent to these two propositions? There are two reasons as far as I can see. The first is a result of Euthyphro’s recognition of the implicit positive argument I have just been discussing, namely, that something being thought to be pious is reason to love it. Therefore, Euthyphro is ready to accept (A).\textsuperscript{50}

Why Euthyphro accepts (B) is a bit murkier. Given the discussion thus far, we might expect Socrates to ask Euthyphro to accept the proposition (C),

\begin{align*}
C: & \sim(\text{The pious is loved by all the gods (}\alpha\text{) because it is being loved by all the gods (}\pi\text{)})
\end{align*}

---

\textsuperscript{49} Euthyphro actually acknowledged reasons for action among the gods earlier at 7e when he agrees that the gods love what they think is beautiful, good, and just. See Allen, \textit{Plato’s Euthyphro}, 45.

\textsuperscript{50} Allen makes a similar point when he says of Euthyphro, “the goodness in things is for him, as for most plain men, a reason for loving and not a consequence of it” (45).
thereby recognizing that the counterpart property of being loved is not an explanation of the gods loving piety. Instead, Euthyphro assents to the claim that it is not the case that the pious is pious because it is loved by all the gods. To understand why Euthyphro assents to (B) we have to do a bit more work.

First, Euthyphro has already agreed that something’s being pious is a reason to love it. If Euthyphro accepted the negation of (B),

D: (The pious is pious ($\pi$) because it is loved by all the gods ($\alpha$))

Euthyphro in effect would be accepting the claim that something being pious is a reason for action for the gods for no other ‘reason’ than that piety is something loved by them. The explanation and justification for the gods loving piety are reduced to their simple subjective preferences. In effect, to say that the pious is pious because it is loved by the gods is equivalent to saying that an action being pious is a reason for action because that action just happens to be loved by the gods. But this is just to reject the very idea of a reason for action, and instead to resort to trying to explain and justify all actions by appealing to counterpart properties.

Once we understand to what Euthyphro assents and why, the rest of Socrates’ refutation of Euthyphro is straightforward. Euthyphro has attempted to define the pious as that which is loved and dear to all the gods, as what is god-beloved. Socrates proceeds to show that the god-beloved and the pious cannot be the same.$^{51}$

---

$^{51}$ Allen provides a less explicit version of this argument. See Allen, *Plato’s Euthyphro*, 43.
Euthyphro’s mistake is to assume that the pious is what all the gods love (6) and that the gods love the pious because it is pious (1), i.e., to accept that the gods act for reasons. The conjunction of these assumptions leads Euthyphro into Socrates’ refutation. But the refutation only succeeds if Euthyphro insists on maintaining the claim that the gods love the pious because it is pious (1), i.e., the gods love what they love because what they love has some further feature that inspires their love, such as its being beautiful, good, or just. Euthyphro must give up either the idea that being pious is a counterpart property (3, 6 and 8), or the idea that the gods act for reasons. He chooses the latter.

The reason the argument works is that Euthyphro does not recognize, while Socrates does, that the word ‘because’ is used in different senses in premises (3) and (5). The ‘because’ of the (1), (2), and (5) is intended to indicate that what is to be

---

52 See Allen, Plato’s Euthyphro, 44-45.
provided is a reason for action, an explanation or justification. In premise (1), that reason for action is accepted, while in premises (2) and (5), that reason is rejected. However, in premise (3), the ‘because’ is intended to indicate why something is defined the way it is, or why we call something what we call it. Therefore, premise (3) is equivalent to ‘those things that are god-beloved are called god-beloved because they are loved by all the gods.’ To put it slightly differently, it is true that the god-beloved is loved, that is after all why we call it god-beloved, but that it is so-loved is not the reason it is loved. The key to understanding the argument is to understand why premise (3) is true but does not provide a reason for action. What we are left with is the idea that being loved by the gods is not the nature or essence of piety, but is rather a result of piety being what it is. Euthyphro has confused the essence of piety with a non-essential property of piety, which serves as a legitimate means of identifying piety.53

The conclusion of the second refutation is that Euthyphro can provide no explanation for what counts as pious and what does not. All he can provide is one reason why what is called ‘god-beloved’ is so-called.

By introducing the distinction between genus and species54, Socrates attempts to get Euthyphro to recognize the fact that what the gods love about pious actions must be some further feature these actions possess independently of being loved by the gods – such as, they are just actions.55

53 While I have put the point differently and I am concerned with reasons for action, see Allen, Plato’s Euthyphro, 45.

54 See Plato Euthyphro 12b.

55 See Plato Euthyphro 12d.
The account of virtue we can extract from the *Euthyphro* is one in which virtue involves a kind of recognition of reasons for action. The *Euthyphro* fails to identify what these reasons for action are, or even if there are any – but what it does establish is that for piety to count as a virtue there must be such reasons, and in order for an individual to qualify as pious, he or she must recognize these reasons. Since identifying reasons for action is central to piety, Plato is committed to a quest for the sources of knowledge that will allow one to have the required ability to recognize appropriate reasons for action.

Whatever the source of knowledge might be, it is clear from Socrates’ reply to Cebes in the *Phaedo* that recognizing reasons for action is recognizing what is best.\(^{56}\) During his reply to Cebes, Socrates describes his disappointment with the teachings of Anaxagoras. Anaxagoras had promised to explain the causes of generation and destruction in terms of the actions of Divine Mind and what is best. Instead, Anaxagoras provided a discussion of cause in mechanistic and physical terms. Socrates’ disappointment is due in part to Anaxagoras’ failure to distinguish physical causes from reasons for action. Reasons for action, or the true and real causes of generation and destruction, must be made in terms of what seems best to the Divine Mind. If Anaxagoras were to fulfill his promise, he would have provided an explanation of the causes and maintenance of the universe in terms of what is best. He does not do this, and since Socrates cannot either, though he thinks it is the best way to understand the concept of cause, Socrates develops his own method for understanding the concept of cause, the theory of forms.

\(^{56}\) See Plato *Phaedo* 95a-100b, particularly 97b-99d.
The Form of the Good is introduced in the *Republic* as an object of knowledge for the guardians. The function of the rational part of the soul is to reason well about the good of the whole soul and to discover the ends worth pursuing. Plato insists that the just man – the man who is able to identify appropriate reasons for action and who acts on the basis of these reasons – can only achieve this status when he has knowledge of the Form of the Good. Plato is arguing (at least) that without a concept of the Good – or at least allowing for or requiring a concept of the Good – there is no way to provide reasons for action. In this sense he is attempting to return to the ‘best way’ of understanding the concept of cause that he hoped Anaxagoras was going to supply in his books.

A reason for action is a reason insofar as it seems best or good to the agent. The reason for pursuing one end instead of another is that this end is ultimately related to some state of affairs that is considered good without qualification, the best. I take it that Plato is arguing that a conception of the Good is required in order to coherently speak of reasons for action. Without a coherent way to talk about reasons for action, there is no coherent way to talk about virtue, since virtue includes the ability to identify reasons for action.

Put another way, if we do not allow for a concept of the Good, then we must admit that every normative description is a counterpart property resulting from our own actions. If every normative description is a counterpart property resulting from our own actions, then normative descriptions provide no reason for action. If normative descriptions provide no reason for action, then regarding something as beautiful, good, or just does not provide reason for action.
2.2.1 Objections and Replies

One might object that the argument that I have provided in my interpretation of the *Euthyphro* does not support the claim that in order for an act of piety to be done piously, the agent must recognize and act upon reasons for action, in particular, the further features of the action and their relationship to certain ends understood as good or best. It may be the case that the gods love pious actions for one set of reasons while in order to act piously an agent must recognize and act upon an entirely different set of reasons.

Suppose Euthyphro had succeeded in providing an account of piety in terms of justice. In that case, one might argue that I am correct that the gods love pious actions because those actions are just. However, in order for an agent to act piously, it might be the case that he or she must act on the basis of no other reasons than that the action is required by the gods, or even that the act is loved by the gods. Knowledge of the gods’ love of a kind of action may be a sufficient reason for pious actions.

Therefore, when Socrates asked, “What is piety?” Euthyphro should not have agreed to (B), but rather asserted and defended its negations: The pious *is* pious because all the gods love it. Their love of pious actions provides human beings with a reason for action. This claim is perfectly consistent with the claim that the gods love the pious because it is pious. The gods have a reason for their love of “pious actions” beyond mere divine subjective preference, and humans have reason to perform “pious actions” beyond mere human subjective preference. The gods’
reasons are that the actions are a part of justice, and human beings’ reasons are that the actions are loved by the gods.

On this understanding of the relationship between reasons for action and piety, Euthyphro’s mistake is not to assume that the pious is what all the gods love and that the gods love the pious because it is pious. Instead, Euthyphro’s mistake is to assume that the way Socrates’ has been using ‘piety’ has a univocal meaning. Therefore, all the argument shows is that piety has one meaning with respect to the gods and another meaning with respect to human beings. That is, with respect to the gods, it is true that the pious is loved because it is pious, and, with respect to human beings, it is true that the pious is pious because it is loved by the gods.

I think this kind of objection can be answered with a refinement of the argument I have presented above. Let pious\(_1\) be “what humans call pious.” Let pious\(_2\) be “what the gods recognize as just.” Then propositions (A), (B) and (D) can be rewritten as follows:

A’: (The pious\(_1\) or \(2\) is loved by all the gods because it is pious\(_2\).)
B’: ~(The pious\(_1\) is pious\(_2\) because it (pious\(_2\)) is loved by the gods.)
D’: (The pious\(_2\) is pious\(_1\) because it (pious\(_2\)) is loved by the gods.)

Euthyphro can consistently agree to both (B’) and (D’). Furthermore, Euthyphro’s definition of piety as god-beloved is acceptable if the definition is confined to piety\(_1\).

This clarification of the argument should help in making the case that humans must act for reasons beyond mere subjective preference, and that ultimately, virtuous activity is never grounded in a mere subjective preference. Pious activity is ultimately grounded in something about the actions called pious which the gods
recognize as good, even if the human agent does not. Moreover, since piety is here being used equivocally, we should distinguish acts of piety from fidelity. One way to interpret the situation with which the objection is concerned is to say that an agent performs acts of piety from fidelity. Whatever virtuous activity this description picks out, it should be distinguished from the acts of piety that are performed for the same reason that the gods love pious acts. Perhaps only the latter is, properly speaking, pious activity.

Nonetheless, the former activity can still be understood as virtuous activity provided the agent understands and acts from an understanding of the genuine good that is promoted and maintained through acts of fidelity to the gods. Obedience, when understood as contributing to and maintaining a genuine good, can be a virtue and not merely the result of manipulation or oppression.

This understanding of the proposed counterexample I think answers the concern expressed in the objection while being consistent with my original interpretation of the *Euthyphro*.

In Section 2.2, I argued that Socrates’ critique of Euthyphro’s second definition of piety reveals something important about the concept of a virtue. An essential feature of the concept of a virtue is that without a conception of the good, the concept of a virtue is reduced to acting from subjective preference rather than for reasons. And while the absence of a substantive account of the Good in the *Republic* diminishes the ultimate effectiveness of Plato’s theory, it does provide us with reason to agree that a virtue, understood as recognizing reasons for action, requires an agent to have a corresponding conception of the good. Euthyphro helps Socrates reveal
these propositions by mistakenly defending a definition of piety that provides no reason for action, initially thinking that being pious was a counterpart property, and confusing the essence of piety with a non-essential property of piety, which serves as a means of identifying it.

2.3 The Rejection of Non-teleological Agent-Based Virtue Theories

There are contemporary varieties of virtue ethics that attempt to deny the claim that a coherent conception of virtue requires a prior understanding of the good and the best.57 Michael Slote offers a useful description of one such variety of virtue ethics, a view often referred to as agent-based virtue ethics.58 Slote is perhaps the most prominent advocate of an agent-based virtue ethics.59 A defining feature of agent-based virtue ethics is the assertion that virtue ethics “must derive its evaluation of human action, whether aretaic or deontic, from independent and fundamental aretaic characteristics of the inner traits or motives of the individual or individuals themselves.”60 So, for example, in Morals from Motives, Slote defends the position that certain motives and traits, such as caring and compassion, have positive moral value and that all act evaluations are derived from prior agent evaluations.61 Slote

57 Christine Swanton couches her objection to Slote in terms of what she sees as his identification of what is right and what is good. This identification is implicit, I think, in describing Slote’s view as a non-teleological agent-based theory. See Christine Swanton, “A Virtue Ethical Account of Right Action,” Ethics 112 (October 2001), 36.


61 See Michael Slote, Morals from Motives.
glosses his description of agent-based virtue ethics by noting how radical a position it is and saying, “it entails that the agent and her inner life are not fixed on any external or independent action-governing moral standard, but rather [the agent and her inner life] constitute the basis and measure of all moral activity.”

While it is true according to Slote that “the admirability of traits and motives are taken to be ethically fundamental” and that moral evaluations of actions are derived from evaluations of character traits, Slote denies that agent-basing entails that acting from virtue is acting in isolation of facts concerning the world. For example, on his view, a properly motivated agent would seek out relevant facts before acting, thereby escaping any possibility of being culpably ignorant. Slote believes he must deny the view that the virtuous agent would act in isolation from facts concerning the world in order to escape the action-guiding objection. In order to act on an intrinsically good motive, an agent must focus on the world as it is and gather facts about the world. Because motives are outward looking in this way, they have the resources to guide action.

However, the action guidance involved in Slote’s response is strictly instrumental in nature. By looking to the world, the agent is seeking information to enable him or her to discover the instrumental means to bring about the achievement of an antecedently accepted end. On Slote’s view all practical reasoning becomes instrumental reasoning. Furthermore, since the source of all value is the agent and his

---


63 See Chapter One above.
or her inner life, facts about the world or about human nature will not reveal anything to an agent about what is valuable, worthwhile, or an appropriate end to be pursued.

Stohr and Wellman have argued that Slote’s position requires that we evaluate actions in a manner that allows terrible actions performed from a good motive to have positive value, or at least no negative value. Stohr and Wellman’s objection begins with the idea that Slote appears committed to the view that an agent who acts from a fundamentally good motive “could not act wrongly no matter what she does.” The concern that Stohr and Wellman present centers on the idea that it is extremely counter-intuitive to suggest that an act is right if it is properly motivated, no matter how much harm of any kind the agent causes. Likewise, it appears extremely counterintuitive to claim that an act is wrong if it is improperly motivated, no matter how much good of any kind is done by the agent. Since there is no source of ethical value beyond the motives of the agent, it appears impossible to say that it would have been ethically better for a properly motivated agent to have acted differently. While I agree that Slote’s agent-basing has the flaws Stohr and Wellman mention, I think the reason for these flaws with agent-basing is that Slote has made the same mistakes Euthyphro made. By defending a definition of right action that provides no reason


65 Stohr and Wellman, “Recent Work on Virtue Ethics,” 52. The initial counter example Stohr and Wellman give is misconceived. The claim is that if an agent killed thousands of individuals out of compassion, then she would not act wrongly. However, it seems to me that it is impossible to kill thousands out of compassion. To claim that one’s motivation in killing thousands was compassion seems to be an indication that one does not understand the word ‘compassion.’ Furthermore, it assumes that the only option available to mitigate the suffering of these thousand individuals was to kill them, an assumption that is almost certainly false.

66 A properly motivated agent would not be culpably ignorant.
for action, Slote is committed to the view that ‘being right’ is a counterpart property and has confused a non-essential feature of right action with its essence.

Another prominent advocate of a form of agent-basing is Linda Zagzebski. Her version of agent-basing is what she calls a pure virtue theory, a theory in which the primary object of normative evaluation is some part of a person’s character. “In a pure virtue theory the concept of a right act is defined in terms of the concept of a virtue or a component of virtue such as a motivation…. [A]n act would not have been right or wrong if it were not for its relation to certain inner personal traits.”

Zagzebski develops a form of agent-based virtue ethics that she calls a motivation-based pure virtue theory in which “the concept of a virtue [is] constructed out of the concept of a good motivation” where the goodness of a motivation is fundamental and is revealed by experience. Motivation is defined in terms of a motive. “A ‘motive’ . . . is an emotion or feeling that initiates and directs action towards an end.” When a certain motive tends to exert a persistent impulse to initiate and direct action towards an end, the motive becomes a motivation. “A motivation [is] a persistent tendency to be moved by a motive of a certain kind.” By substitution, a motivation is a persistent tendency to be moved by a certain kind of emotion or

---

67 Zagzebski, *Virtues of the Mind*, 79. She also says, “By a pure virtue theory…I mean a theory that treats act evaluation as derivative from the character of an agent. Roughly, an act is right because it is what a virtuous person might do” (Zagzebski, *Virtues of the Mind*, 16).

68 In “An Agent-Based Approach to the Problem of Evil,” Zagzebski gives an account of the moral experience which she believes indicates that some motivations are ethically fundamental. We sometimes meet, she argues, individuals whose “goodness simply shines forth from the depths of their souls” (130). When we attempt to discover what accounts for this phenomenon, the available explanation, she argues, is that we recognize certain inner traits as good in themselves and not for the sake of anything independently identified as good. See also Zagzebski, *Virtues of the Mind*, 83-84.

69 Zagzebski, *Virtues of the Mind*, 131.

70 Zagzebski, *Virtues of the Mind*, 132.
feeling that initiates and directs action towards an end. A motivation, as a persistent
tendency to action, is more firmly entrenched among an agent’s inner personal traits
than is a motive. It is this feature of a motivation that allows motivation to play such
a central role in a pure virtue theory. “A virtue, then, can be defined as a deep and
enduring acquired excellence of a person, involving a characteristic motivation to
produce a certain desired end and reliable success in bringing about that end.”

Or, by substitution, a virtue is a deep and enduring acquired excellence of a person,
involving a characteristic persistent tendency to be moved by a certain kind of
emotion or feeling that initiates and directs action towards a certain desired end and
reliable success in bringing about that end.

The normative quality of the desired end in Zagzebski’s definition is
ambiguous. Given the way the definition is presented, one can reasonably ask
whether a virtue can involve a motivation for any desired end or whether a virtue can
only involve a motivation for a desirable end, that is, an end that ought to be desired.
This ambiguity is easily resolved, though not in Zagzebski’s favor, when we recall
that Zagzebski said earlier that she would define virtue in terms of a good motivation
where the goodness of the motivation was fundamental and revealed by experience.
Therefore, Zagzebski appears to be equivocating between desired and desirable.
Since her view is an agent-based theory, and therefore a non-teleological version of
virtue ethics, she cannot appeal to some end or good which ought to be desired
because it is good. Instead, her only recourse is to insist that the ends actually

---

71 Zagzebski, Virtues of the Mind, 137.
72 In her discussion of types of virtue theories in Virtues of the Mind, Zagzebski makes a
number of distinctions which parallel Slote’s distinctions. For example, what Slote calls an agent-prior
desired are desirable because they are the objects of desire identified by motives which are good independently of the ends they desire. It is this one feature of agent-basing to which I will object during the discussion that follows.

To complete her description of agent-basing, Zagzebski claims that to have a virtue means that one has reliable success in bringing about the end toward which the virtue aims. Therefore:

An act \( A \) is an act of virtue \( V \) iff,

1) \( A \) arises from the motivational component of \( V \),
2) \( A \) is an act that persons with virtue \( V \) characteristically do in the circumstances surrounding \( A \),
3) \( A \) is successful in bringing about the end \( E \) of virtue \( V \), and
4) \( (3) \) “because of”\(^{74} \) (1) and (2).

view, Zagzebski calls a good-based view. In addition, Zagzebski sometimes calls good-based views, especially a form of a good-based view she calls happiness-based, teleological virtue ethics. By contrast, agent-based views, of which a motivation-based view is a form, she sometimes refers to as a non-teleological virtue ethics. It is essential to understand that for Zagzebski the distinction between a teleological and a non-teleological view is a distinction concerning the relationship between virtue and good, and not a distinction concerning the relationship between virtue and right action. In “An Agent-Based Approach to the Problem of Evil,” Zagzebski makes the distinction between a teleological and a non-teleological pure virtue ethics this way: “Common teleological forms [of pure virtue theories] make the concept of a good life the fundamental ethical concept and a virtue is explicated in terms of its contribution to a good life, either as a means to it or as a constituent of it…. A more radical, non-teleological form of virtue ethics makes the virtue of other or other internal properties of the agent ethically fundamental, and the good is treated as a derivative concept. This is what I am calling agent-based ethics” (129).

\(^{73}\) These conditions are adapted from an unpublished draft of an manuscript by Linda Zagzebski, which was published as “What is Knowledge?” in The Blackwell Guide to Epistemology, ed. John Greco and Ernest Sosa (Malden, Mass: Blackwell Publishers, 1999). Hereafter this draft manuscript will be referred to as “What is Knowledge? (Draft).” “What is Knowledge? (Draft)” is consistent with the account in Virtues of the Mind in most respects, although there are slight discrepancies in the language of some of the definitions.

\(^{74}\) Zagzebski admits in “What is Knowledge? (Draft)” that the notion of ‘because of’ is vague and that she has no adequate account of the notion.
In condition (1), the motivational component of V is “a disposition to have an emotion that directs actions towards an end”\textsuperscript{75} where presumably the end in question is the end E aimed at by V, as mentioned in condition (3). Condition (3) guarantees that an act of virtue be successful in achieving its particular end. ‘An act of virtue’ is therefore a success term. If A did not reach its appropriate end E, A would not be an act of virtue, although Zagzebski suggests that it may very well be an act of a virtuous person. Possessing a particular virtue, even in a deeply entrenched sense, does not guarantee that all of one’s relevant actions will be acts of that virtue. Also, it is not necessary on this definition that a person possess virtue V in order to perform an act of virtue V. But this definition of an act of virtue does require the agent to have the motivation associated with the virtue, that is, an agent must have a persistent tendency to be moved by a certain kind of emotion or feeling that initiates and directs action towards the end of V.

Recall, however, that Zagzebski’s account is a non-teleological view. As a non-teleological view, the concept of the good, and therefore the desirability of the end associated with V is entirely derived from the goodness of the virtue itself. Therefore, when Zagzebski says “a right act is what a person who is virtuously motivated, and who has the understanding of the particular situation that a virtuous person would have, might do in like circumstances”\textsuperscript{76} and “a right act, all things considered, is what a person with phronesis might do in like circumstances”\textsuperscript{77} she is making no appeal to the good of the end sought by the action in order to establish its

\textsuperscript{75} Zagzebski, “What is Knowledge? (Draft),” 29.

\textsuperscript{76} Zagzebski, \textit{Virtues of the Mind}, 235.

\textsuperscript{77} Zagzebski, \textit{Virtues of the Mind}, 239.
rightness. Instead, “the concept of a right act is strictly derived from the concept of a virtue. The motivation or behavior of virtuous persons is what makes an act right.”

Let us call accounts of agent-basing such as Slote’s and Zagzebski’s ‘non-teleological agent-based virtue ethics.’ Justin Oakley initially describes all virtue ethics as non-teleological when he claims that all virtue theorists are committed to the view that “an action is right if and only if it is what an agent with virtuous character would do in the circumstances.” He glosses this description by saying that “a right action is one that is in accordance with what a virtuous person would do in the circumstances, and what makes the action right is that it is what a person with a virtuous character would do here.” On its face, Oakley’s description appears to be a description of non-teleological agent-based virtue ethics. I will argue that any version of virtue ethics that incorporates the view Slote and Oakley describe is incoherent. The first clause of Oakley’s description is true in the same sense that Euthyphro’s second definition is acceptable when applied to piety. However, there is no sense in which the second clause of Oakley’s description is correct. I will argue that it cannot be correct for the same reasons that Euthyphro’s second definition is unacceptable when applied to piety; it explains and justifies an action by appealing to a counterpart property.

78 Zagzebski, “An Agent-based Approach to the Problem of Evil,” 129. (Emphasis is original.)


80 Oakley, “Varieties of Virtue Ethics,” 130. (Emphasis is original.)

81 He does provide a further discussion of what this condition means, but at the moment, I want to focus on the natural interpretation of his initial claim.
Recall that according to Socrates’ argument refuting Euthyphro’s second definition of piety, a counterpart property can never be given as an explanation or justification of the actions of an agent. Since agent-based virtue theorists define right action in terms of what a virtuous agent does, ‘being right’ is a counterpart property of the actions of virtuous agents. Therefore, just as

M: \neg (\text{Something affects } X (\alpha) \because X \text{ has the property ‘being affected’} (\pi))
N: (X \text{ has the property ‘being affected’} (\pi) \because \text{something affects } X (\pi))

so too

S: \neg (X \text{ is performed by a virtuous agent} (\alpha) \because X \text{ has the property ‘being an action a virtuous person would perform’} (\pi))
T: (X \text{ has the property ‘being an action a virtuous person would perform’} (\pi) \because \text{a virtuous agent would perform it} (\alpha))

Definition: let X equal right action

X: \neg (\text{Right action is performed by a virtuous agent} (\alpha) \because \text{right action has the property ‘being an action a virtuous person would perform’} (\pi))
Y: (\text{Right action has the property ‘being an action a virtuous person would perform’} (\pi) \because \text{a virtuous agent would perform it} (\alpha))

Recall that the point of this kind of distinction in the *Euthyphro* was to make the negative point that the actions of an agent are conditionally prior to the counterpart properties of a patient. Therefore, the counterpart property can never be given as an explanation or justification of the actions of a virtuous agent. Recall also that the
positive implication of Socrates’ argument is that an explanation of or justification for action must provide a reason for action. However, \( (Y) \) does not provide a reason for action. It merely provides an explanation for why an action has a particular property. Therefore, we are left wondering why a virtuous agent performs the actions she does. Presumably, she would perform an action because she regards the action as good, correct, appropriate, \( etc. \), or because she regards them as right. This is the kind of common sense proposition to which Euthyphro agreed:

\[
W: \text{(Right action would be performed by a virtuous agent (}\alpha\text{) because it is right (}\pi\text{))}
\]

However, the agent-based virtue theorist wants to assert:

\[
V: \text{(Right action is right (}\pi\text{) because a virtuous agent would perform it (}\alpha\text{))}
\]

Recall that Euthyphro rejected a statement of the form of \( (V) \) in favor of its negation. And he was correct to do so. Accepting \( (V) \) would be accepting the claim that an action being right is a reason for action for the virtuous agent for no other ‘reason’ than that right action is something he or she happens to do. But this is just to reject the very idea of a reason for action, and instead to resort to trying to explain and justify all actions by appeal to counterpart properties.

The virtuous agent appears to be in no position to answer the question ‘Why do you regard this action as right, correct, appropriate, \( etc. \)?’ Her only available answer is that she just does so regard it. What makes the action right, correct, appropriate, \( etc. \), is that it is the kind of action she would perform in the
circumstances. If she has no alternative reason for explaining why she thinks that her action is right, correct, appropriate, etc., on what basis can we deny that the virtuous agent acts on no other basis than her mere preference? If the virtuous agent acts based on mere preference, on what grounds can her actions be distinguished from the non-virtuous agent’s actions? Recall that on a non-teleological agent-based theory, appeal to the character of the virtuous agent explains and justifies right action. But now we find out that the character of the virtuous agent is such that she acts on the basis of mere preference. If acting based on mere preference is what distinguishes the virtuous agent from the non-virtuous agent, then what use is the concept of virtue in distinguishing right action from wrong action? Indeed, the appeal to an agent-based notion of right action allows for the possibility that two virtuous agents need not even regard the actions of the other as acceptable.

The agent-based virtue theorist faces a choice similar to Euthyphro’s. He or she must either abandon the belief that ‘being right’ is a counterpart property of the actions of the virtuous, or give up the belief that virtuous agents act for reasons. Euthyphro opted for the latter. I suggest the agent-based virtue theorist opt for the latter as well.

The mistakes of the agent-based virtue theorist are the same mistakes Euthyphro made. He or she defends a definition of right action that provides no reason for action, believes that ‘being right’ is a counterpart property, and has confused the essence of being right with a non-essential feature, which can be used to identify right action.
Now, only on the agent-based theory is ‘being right’ a counterpart property of the actions of an agent. On non-agent-based theories (or act-based theory), the definition of right is made in terms of some good independent of the virtuous agent. It is the virtuous agent’s recognition of a particular action’s relationship to this independent good that provides him or her with a reason for action.\textsuperscript{82} On non-agent-based theories ‘being right’ is not a counterpart property of an agent’s action, but rather it constitutes a reason for action. Whether its being right is ethically fundamental or not is another question that would need to be answered by any non-agent-based theory.\textsuperscript{83} However that question is ultimately resolved, we can at least be confident that right action cannot be derived from the actions of any agent.

If (V) is true, it is only in the following sense: Let right\textsubscript{1} be “what humans call right” and let right\textsubscript{2} be “has the appropriate relationship to an agent independent good or end.” Then,

\[
V': \text{(Right}_2 \text{ action is right}_1 (\pi) \text{ because a virtuous agent would perform it (}\alpha))
\]

(V') is true in the sense that we call actions right because we know that virtuous agents recognize, while we may not, that there are reasons to perform those actions. However,

\textsuperscript{82} Michael Slote describes an act-based virtue theory as one that holds that “being virtuous involves being keyed into facts independent of one’s own virtuousness about what acts are admirable or called for” (Slote, “Virtue Ethics” in Three Varieties of Ethics, 178).

\textsuperscript{83} That is, a task of virtue ethics, in addition to giving an account of the moral psychology of virtue and vice, is to provide an account of the good, or at least to allow some minimal conception of the good to operate centrally in the moral psychology of the virtues.
V": (Right$_2$ action is right$_2$ (π) because a virtuous agent would perform it (α))

is false because the action’s appropriate relationship to the good cannot be a result of a particular agent having performed it. It cannot be the result of a counterpart property.

2.3.1 Objections and Replies

It might be suggested that my critique has no force because the preferences of the virtuous agent are not ‘mere preferences’ but the preferences of a virtuous agent and are therefore accorded a special status. While it is true that the preferences of non-virtuous agents do not provide justifiable reasons for action, the preferences of the virtuous agent do provide justifiable reasons for action – the reason is simply that an agent with virtuous preferences prefers them.

Surely, this objection is inadequate. On the non-teleological agent-based view, the internal states of the agent are fundamental. However, everyone has internal states. How are the internal states of the virtuous agent to be distinguished from the internal states of the non-virtuous? The agent-based virtue theorist cannot appeal to something external to the agent in order to make this distinction for to do so would be to admit that some feature of the world external to the virtuous agent is fundamental and hence would be a rejection of agent-based accounts of virtue ethics.\textsuperscript{84} Without an appeal to something other than the internal states of the virtuous

agent, the agent-based virtue theorist has failed to give any analysis of the concept of virtue, of the virtuous agent, or of right action.

One might respond to this reply by weakening the claims of the agent-based theory, arguing that the agent-based theory is only committed to the view that the internal states are primary in the explanation of virtuous behavior, not that the internal states of the virtuous agent make an action right; internal states are not primary in the justification of virtuous behavior. This reply attempts to avoid the force of the argument by making a distinction between making an action right and explaining why an action was done. However, if human actions are explained by appealing to reasons for action, and the virtuous agent is able to identify the features of an action that make it right and regard these as reasons for action, then there is no difference between explaining the actions of a virtuous agent and describing what makes an action right.

2.4 The Rejection of Teleological Agent-based Virtue Ethics

Agent-based virtue theory is the claim that what makes an action right is that it is what a person with a virtuous character would do in those circumstances. Above I have argued that there is a deep misunderstanding embedded in any non-teleological virtue theory that takes this view to be correct. Some versions of agent-based virtue ethics do not feel compelled to make the stark choice I recommended to the non-teleological agent-based virtue theorist. Recall that I suggested that the non-teleological agent-based virtue theorist faces a choice similar to Euthyphro’s. He or she must either abandon the belief that ‘being right’ is a counterpart property of the actions of the virtuous agent, or give up the belief that virtuous agents act for reasons.
Some versions of agent-based virtue ethics are more sophisticated in that they attempt to ground the value of the virtues in something external to the agent, and therefore appear to escape the argument I have deployed against non-teleological agent-based virtue ethics. There are two general strategies of Aristotelian inspiration for developing a teleological agent-based virtue ethics. One Aristotelian approach grounds the value of the virtues in the concept of a good human being. Another common Aristotelian approach grounds the value of the virtues in what is good for human beings. These views are capable of stating that what makes an action right is that it is what a virtuous agent would do in the circumstances and that what makes the virtuous agent the standard is that his or her character traits are either features of a good human being or are features that are good for humans to have. Since on this view the value of the virtues is derivative, and the notion of the good is fundamental, this view is a teleological virtue theory. However, it is also an agent-based theory because it defines right action in terms of the virtuous agent. Therefore I will be referring to these views as a teleological agent-based virtue theory.

Rosalind Hursthouse, in her earlier work, takes the first Aristotelian approach toward grounding the values of the virtues in the concept of a good human being. When responding to the objection that virtue ethics is hopelessly circular because it defines right action in terms of the virtues and the virtues in terms of right action, she

---

85 In On Virtue Ethics, (Oxford: Oxford UP, 1999), Rosalind Hursthouse attempts to combine both these approaches into a single Aristotelian approach. In “A Virtue Ethical Account of Right Action,” Christine Swanton refers to Hursthouse’s view as a “qualified agent’ account of rightness” in which the standard or criteria for rightness is determined by appeal to an agent who is qualified to make determinations of rightness due to his or her status as a virtuous agent.

argues that such an objection is misconceived because Aristotelian virtue ethics does not merely define right action in terms of the virtues, but goes on to define the virtues in terms of eudaimonia.\textsuperscript{87} When replying to the objection that Aristotelian virtue ethics is committed to a reduction that involves defining all moral concepts in terms of virtuous agents, she argues that the moral concepts of virtue ethics are dependent on the concepts of good and evil, harm and benefit, worthiness and worthlessness, advantageous and detrimental, and pleasant and unpleasant.\textsuperscript{88} According to Hursthouse, Aristotelian virtue ethics is misunderstood when considered to be reductive or non-teleological. All Aristotelian approaches to virtue ethics appeal at some point to the concept of the good. As she rightly points out, it would be absurd to take moral advice from someone of unquestionable motives but admittedly ignorant about what is truly good.\textsuperscript{89}

Michael Slote presents an argument against the teleological agent-based virtue ethics of Hursthouse.\textsuperscript{90} According to Slote, on Hursthouse’s view “acts are said to be right if they accord with or exercise the virtues, and a trait is said to be a virtue if people need it in order to flourish and have good lives.”\textsuperscript{91} Slote argues, I think rightly, that Hursthouse cannot provide an explanation for why the criteria for a trait being a virtue is not also the criteria for an action’s being right.

\textsuperscript{87} Hursthouse, “Virtue Theory and Abortion,” 220.

\textsuperscript{88} See Hursthouse, “Virtue Theory and Abortion” and Hursthouse, “Applying Virtue Ethics.”

\textsuperscript{89} Hursthouse, “Applying Virtue Ethics,” 74.

\textsuperscript{90} See Slote, “Virtue Ethics,” in \textit{Three Methods of Ethics}. He characterizes Hursthouse’s view as a happiness-based agent-prior account.

\textsuperscript{91} Slote, “Virtue Ethics,” in \textit{Three Methods of Ethics}, 207.
In section 2.4.1 I will be arguing that Hursthouse’s attempt to provide a teleological agent-based virtue ethics is conceptually confused in ways similar to the way in which non-teleological agent-based virtue ethics is conceptually confused. I hope to show in sections 2.4.2 and 2.4.3 that Hursthouse’s appeal to agent-basing is also redundant and should be rejected.

2.4.1 Hursthouse as a Teleological Agent-based Virtue Theorist

The following characterization of right action is a consistent feature in Hursthouse’s work.\textsuperscript{92}

1. An action is right if and only if it is what a virtuous agent would characteristically do in the circumstances.
2. A virtuous agent is one who acts virtuously, that is, one who has and exercises certain character traits, namely, the virtues.
3. A virtue is a character trait that … (a human being needs for \textit{eudaimonia}, to flourish or live well).

With these principles Hursthouse wants to make explicit the conceptual links between right action and the virtuous agent, the virtuous agent and the virtues, and virtues and flourishing. In “Virtue Ethics and Human Nature,” Hursthouse offers a helpful interpretation of this characterization of right action. She suggests that the appropriate way to understand these three conditions is as answers to three prior questions: What is a right action? A right action is an action a virtuous agent would characteristically perform in the circumstances. What is a virtuous agent? A virtuous

agent is an agent who has and exercises the virtues. What are the virtues? The virtues are those character traits that a human being needs for *eudaimonia*.

In *On Virtue Ethics*, Hursthouse deploys this characterization of right action while responding to claims that characterizing right action as ‘what the virtuous agent would do’ is circular. The argument for circularity depends on the assumption that the only way a virtue theorist can define the virtuous agent is as an agent that performs right action. She attempts to escape the charge of circularity by asserting that she is not trying to derive ‘right’ from the concept ‘virtuous agent.’ She also denies that the only way to give an account of the virtuous agent is in terms of right action. Instead, she simply argues that the string of conceptual links beginning with right action ends with the flourishing. Furthermore, she admits that the initial specification of right action in (1) is uninformative. It is only after the further specifications in (2) and (3) are made that (1) becomes informative. Before these specifications are made, (1) is, strictly speaking, empty. (2) is also, strictly speaking, empty without the specification provided in (3). Which traits count as virtues is yet to be determined. One option for determining which traits count as virtues is to provide an abstract test; that is what the parenthetical clause in (3) provides.

At this point, let me introduce a distinction. Virtue\(_n\) is a normative term while virtue\(_p\) is a psychological term. Virtue\(_n\) or virtuous\(_n\) refers to actions and their relationship to the good.\(^{93}\) An action being virtuous\(_n\) does *make* an action right.\(^{94}\)

\(^{93}\) I have chosen to speak of virtue\(_n\) in terms of a relationship to the good rather than in more neutral terms such as ‘the targets’ toward which a virtue aims. The language of ‘targets’ appears to have originated with Robert Audi, *Moral Knowledge and Ethical Character* (New York: Oxford UP, 1997), 180, and has subsequently been adopted by Christine Swanton in “A Virtue Ethical Account of Right Action.”
However, being virtue\(_n\) is the result of a particular action’s relationship to the good, not the result of the fact that a particular kind of person would perform it; that is, it is not the result of virtue\(_p\). Virtue\(_p\) refers to a particular kind of mental state, a mental state that allows the agent to consistently recognize objectively appropriate reasons for action as well as to reliably act for, or because of, those reasons. A person with the virtues\(_p\) is virtuous\(_p\). It follows then that (at least some) actions can be virtue\(_n\) without being virtue\(_p\).\(^{95}\) For example, it is possible for an action to be an act of honesty without being an honest action.\(^{96}\) If honesty\(_p\) is understood as the ability to recognize and speak the truth, it is possible for an agent to selectively see the truth, selectively speak the truth, or both. In fact, a common vice\(_p\) is the propensity to selectively speak the truth in such a way that the speaking of the truth is used to inflict harm on others. When such a person says something rude, his or her defense is often to claim that he or she is “just being honest.” And if what he or she is saying is true, there is a sense in which he or she is performing an act of honesty. Of course, he or she is also “just being cruel.” That is, he or she is performing an act of rudeness.

\(^{94}\) This distinction is, I believe, a common one, though perhaps often not made explicit. Christine Swanton attempts to make a similar distinction explicit in “A Virtue Ethical Account of Right Action,” 37-39. Swanton uses the confusing language of virtuous action and action from virtue. A virtuous action is an action that succeeds in “hitting the target of the virtue.” An action is from virtue if the action displays the appropriate excellence of character, or virtue. Furthermore, an action from virtue is not necessarily a virtuous act nor is a virtuous action necessarily an act from virtue. On Swanton’s account, an act is right if it is “overall virtuous.” Therefore, Swanton’s account of right action is not agent-based, but is dependent on the action “hitting the target of the virtue.” See also Swanton, “A Virtue Ethical Account of Right Action,” 45.

\(^{95}\) This distinction is implicit in the distinction between an act of virtue and a virtuous act.

\(^{96}\) I am not using ‘virtuous action’ in the same way Swanton uses it. Her use of ‘virtuous action’ is closer to what I am calling ‘an act of virtue.’ Also, Swanton’s ‘action from virtue’ is closer to what I call ‘virtuous action.’ Despite the discrepancy in vocabulary (and I suspect that the discrepancy may be the result of a third conceptual distinction existing in the neighborhood) Swanton and I agree that “an action can be just or temperate if it hits the target of the virtues of justice or temperance, and an action may hit those targets without exhibiting a just or temperate state” (Swanton, “A Virtue Ethical Account of Right Action,” 37).
In addition, he or she most likely is also performing a rude act and is rude.

Nonetheless, it is appropriate to describe the act of truth telling as honest in one sense of the word while it may be inappropriate in another sense of the word. Since what is said is appropriately related to the good insofar as it is true, what is said is accurately described as honest. In so far as it is not the result of a virtuous character it is not accurately described as honest, but instead is more accurately described as rude.

Virtue, and virtue, are rarely explicitly distinguished. One result of failing to distinguish them is to make the mistake that Hursthouse makes in Chapter 1 of On Virtue Ethics. Making this distinction explicit generates the following list of conditions for right action:

1'. An action is right if and only if it is what a virtuous agent would do in the circumstances.

2'. A virtuous agent is one who acts virtuously, that is, one who has and exercises certain character traits, namely, the virtues.

3'. A virtue is a character trait that … (a human being needs for eudaimonia, to flourish or live well.)

Premise (1') is Hursthouse’s agent-basing premise. Although it is true that Hursthouse introduces the concept of the good in (3'), a strict reading of her conditions shows that the fact that an action is right is the result of the psychological state of the agent who performs the action, at least in those cases in which the agent in question is a virtuous agent. So, for example, assume that an agent, VA, is a virtuous agent. Therefore, by premise (1'), VA’s actions are right because VA performs them.

97 It is true that Hursthouse describes 1' as a “specification” rather than as a definition. Nonetheless, she is attempting to give an account of right action and of the way we should act. Her first move is to present what she calls an “agent-centered” premise. My contention is that this move is in relevant respects equivalent to introducing an agent-based premise. Furthermore, the introduction of 2' and 3' will not allow Hursthouse’s account of right action to escape from the kind of objection I presented against non-teleological agent-based virtue ethics.
The result is that Hursthouse’s account of the right action is an agent-based account. Therefore, Hursthouse is committed to proposition Y".\(^{98}\)

Y" states:

\[ Y" : (VA’s \text{ action is right (} \pi \text{) because he or she, a virtuous agent, performs it (} \alpha \text{))} \]

But our discussion of counterpart properties and conditional priority has shown that such a description of right action is illicit, as it was in the case of D and Y'. Therefore, as a description of the rightness of the actions of a virtuous agent, Hursthouse’s account of right action must be rejected.

Even though Hursthouse appears committed to this flawed account of right action, she also attempts a related but different attempt to escape the original objection that any virtue ethical account of right action will be circular because it attempts to define right action in terms of the virtuous agent and the virtuous agent in terms of right action. The argument for circularity immediately generates the criticism that virtue ethics cannot be action guiding – that it cannot generate rules to guide us toward the performance of right action. Hursthouse insists that virtue ethics does generate action guiding rules (which she calls v-rules), such as, ‘Do what is honest.’ In her description of how v-rules are generated from the common and familiar list of virtues, she is clearly using virtue\(_n\). If Hursthouse is using virtue\(_n\) instead of virtue\(_p\) in order to generate v-rules, then it appears that she is not providing an agent-based account of right action after all.

---

\(^{98}\) Recall that Y stated: (Right action has the property ‘being an action a virtuous person would perform’ (\(\pi\) because a virtuous agent would performs it (\(\alpha\))).
That Hursthouse shifts from using virtue_p to using virtue_n is evident in her method of deriving v-rules from a list of the common and familiar virtues (where she means virtues_n) rather than provide an abstract test for deriving v-rules from what virtues_p are necessary in order to flourish, live well, or achieve eudaimonia. The list of virtues she provides, as she implicitly states in footnote 7, are not lists of virtues_p, but of virtues_n. Hursthouse insists on putting v-rules in terms such as ‘Do what is honest’ rather than expressing them with adverbs, as in ‘Act honestly.’ She does this because “the adverbs connote not only doing what the virtuous agent would do, but also doing it ‘in the way’ she would do it, which includes ‘for the same sort(s) of reason(s).’” Since on her account it now appears to be sufficient for right action to act as the virtuous agent would act, the adverbial construction is unnecessary. It might appear then that Hursthouse can escape the kind of objections I have been leveling against agent-based accounts of right action by appealing to virtue_n instead of virtue_p. However, I do not believe she can escape the objection. On Hursthouse's account, the v-rule 'Do what is honest' is a moral rule because it commands or requires an agent to perform an action that is right, that is required. But the fact that the action is right or required is still derived from the claim that the action is right because it is the sort of thing a virtuous agent would do. Therefore, Hursthouse is committed to proposition Y".  

---


100 Recall that Y stated: (Right action has the property ‘being an action a virtuous person would perform’ (π) because a virtuous agent would perform it (α)).
Y''' states:

Y''': (The action a v-rule requires is right (π) because a virtuous agent would perform the action specified in the v-rule (α))

But our discussion of counterpart properties and conditional priority has shown that such a description of right action is illicit, as it was in the case of D, Y' and Y''. In putting forward her v-rules, Hursthouse has again appealed to a counterpart property as an acceptable justification for an action.

2.4.2 A Teleological Good-based Virtue Theory

What follows from my argument, I think, is that there is no coherent way, even on a teleological agent-based virtue ethics, to make sense of the claim that what makes an action right is that it is what a virtuous agent would do. What also follows, it seems to me, is that there is no way to assess the rightness or wrongness of an action merely on the basis of the virtues had or not had by the agent that performed it.\(^{101}\)

Given this critique of Hursthouse’s teleological agent-based virtue ethics, it appears that a better analysis of right action is one that bypasses the psychological states of the virtuous agent altogether, unless of course the psychology of the agent is directly related to the good, as it sometimes is. In her discussion of the virtuous agent’s reasons for action, Hursthouse notes that the reasons given by a virtuous agent

\(^{101}\) While I have emphasized the priority of knowing the agent’s reason for action in order to assess an agent’s actions, Christine Swanton has also emphasized that Hursthouse’s theory lends itself to serious miscalculations of an agent’s actions because it does not take seriously enough the limitations of human rationality and knowledge. See Swanton, “A Virtue Ethical Account of Right Action,” 35.
for why he or she performed the action in question “all show or indicate what the agent took as relevant or salient, advantageous or disadvantageous, good or evil, decisive or compelling, about the action or the situation or both.”\textsuperscript{102} And, as she also notes, the virtuous agent gets these things correct.

So, in her account of the virtuous agent’s reasons for action, we see that right action has nothing to do with what the virtuous agent might do. The reason an action is right is that it is in accord with what \textit{is} relevant, salient, advantageous, good, decisive, or compelling. This is exactly why Hursthouse has to mention that virtuous agents get things correct. The reasons for right action are often wholly independent of the virtuous agent. That is, what is virtuous\textsubscript{a} can be wholly independent of the virtuous agent. All an appeal to a virtuous agent adds to an account of right action is an example of someone with character who gets things correct.

It seems to me that a better way to forge a conceptual link between the various concepts in which Hursthouse is interested is along the following lines:

1'. An action is right if and only if it has an appropriate relationship to the good.
2'. Having an appropriate relationship to the good makes an action virtuous\textsubscript{a}.
3'. An act being virtuous\textsubscript{a} is consistently and reliably recognized as a reason for action by those possessing the virtues\textsubscript{p}.
4'. Those possessing the virtues\textsubscript{p} are virtuous\textsubscript{p} agents.
5'. Being a virtuous\textsubscript{p} agent is necessary for ...(\textit{eudaimonia}, to flourish or live well).

These premises are merely an initial suggestion. However, notice that they are consistent with either version of the Aristotelian strategy of grounding the value of the virtues in the concept of a good human being or in the concept of what is good.

\textsuperscript{102} Hursthouse, \textit{On Virtue Ethics}, 129.
for human beings. I think it is clear that it is conceptually less confusing to directly
describe right action as acting in accord with the correct reasons. This was Slote’s
point, which I mentioned above. Why go to the effort of defining right action in
terms of the virtuous agent? We can still say that an action should be performed
because it is honest. But its being honest is a feature of its relationship to some
conception of the good – say the necessity of trust in a community. By omitting the
necessity of reference to the agent, we can maintain room for a distinction between
right action and virtuous action. Furthermore, we can provide a framework in which
education can occur. We teach children to perform right actions, explaining along the
way that such actions are related to the good in various ways, rather than simply
stopping at “Don’t lie because it is dishonest.” That is uninformative. Children are
natural philosophers¹⁰³ and are bound to ask what dishonesty is. Hursthouse has
stopped at the list of the virtuesn - but the child and the philosopher want to know
what dishonesty is. Socrates certainly would have wanted to know.

What I think I have achieved is to dispel the apparent need for virtue ethics to
define right action in terms of what a virtuous agent would do in the circumstances.
Recall that Oakley argues that all virtue ethics share the claim that what makes an
action right is that it is what a virtuous agent would do in the circumstances. In a
further explication of the meaning of this condition, he suggests, “virtue ethics holds
that acting out of the appropriate motive and disposition is necessary for right
action.”¹⁰⁴ I am suggesting that there is reason to believe this claim is false. What

¹⁰³ Honer, Hunt, and Okholm, Invitation to Philosophy, (Belmont: Wadsworth, 2002), 2.
may be true is that acting out of the appropriate motive and disposition is necessary for human flourishing. However, it appears to be sufficient for right action that one’s actions have the appropriate relationship to the good regardless of the dispositions or motives from which the agent acts. We describe such actions as acts of virtue, acts of piety, acts of charity, not because the agent is acting from the proper dispositions, and not because they are the sorts of actions that a virtuous person would perform, though that is true; we call them acts of virtue because they have a relationship to the good that is best described in terms of virtue concepts such as piety or charity.

So, what is the general implication of this kind of argument against both non-teleological and teleological agent-based theories? The first implication is that virtue ethics cannot regard something entirely internal to the virtuous agent as ethically fundamental when attempting to provide an account of objectively appropriate reasons for action. The second implication is that any virtue theory, regardless of what it takes to be ethically fundamental, must provide an account of how the virtuous agent perceives reasons for actions derived from that which is ethically fundamental and furthermore, how the virtuous agent acts for those perceived reasons.

2.4.3 Replies and Objections

One might object to my criticism of Hursthouse in the following manner.\(^\text{105}\)
The v-rule ‘do what is honest’ instructs us to do the objectively right thing under a

\(^{105}\) I thank Sabina Lovibond for actually objecting in this manner during her comments on an earlier draft of this chapter presented at the Oxford University Graduate Philosophy Conference, November 17\(^{\text{th}}\), 2002.
certain aspect of good, and so it invokes an appropriate virtue. However, Hursthouse has not equivocated, since the v-rule merely commands us to model our actions on those of the virtuous person to the extent that this is something that can be commanded, i.e., to the extent that it directs our attention to features of the virtuous person’s conduct that we can reproduce at will (in the case of honesty, paying our bills, not lying, etc.). Hursthouse could only be accused of equivocating if she were insisting that the v-rule commands us to direct our attention to something we cannot reproduce at will, namely that entrenched disposition of the will that constitutes honesty in the virtuous person.

One might also object that Hursthouse’s account of right action is of significant epistemic value. One might argue that in some circumstances - say, where one was unsure whether or not a certain action would be right in respect of being honest, or permissible from the point of view of honesty - it might be quite helpful to know that the action was, or was not, the sort of action that an honest person (meaning one who had the entrenched disposition of honesty, or honesty_p) would do. The observation that ‘an action is right if and only if it is what a virtuous person would do in the circumstances’ expresses a thought that is potentially of some epistemic value to individual agents, and moreover a thought that trades - not in a vicious way but in a constructive way - on the fact that one can ‘do what is honest’ without (or without yet) being a (thoroughly) honest person oneself. Sometimes it just is more apparent to one that VA is a virtuous person than that action A would, or

106 Again, I thank Sabina Lovibond for this criticism.
would not, be compatible with honesty, and one can make use of that better known fact in order to advance one’s understanding of what virtue demands.

These two objections are intimately related. Certainly a virtue theory is not merely concerned with having agent’s perform right action (acts of virtue), but also wants agents to perform virtuous actions. So, in response to the first objection that Hursthouse does not equivocate because v-rules are only intended to direct our attention to what we can control at will, namely our immediate actions, I must point out that the rules of a virtue theory must not be directed merely to the direction of action, but also to the education of the agent who receives them. I would agree with Hursthouse that in certain circumstances a v-rule is properly interpreted in a way that does not require an adverbial construction. And when it is so understood, it is often enough to command right action, though not for the reasons Hursthouse claims. On the other hand, commanding right action is not the sole purpose of a v-rule. Rules are a means of education, provided that the rule is promulgated together with an education concerning the point and purpose of the rule. And to educate an agent in the point and purpose of the rule is to provide an account of the good and the best that the observance of the rule is meant to either facilitate or constitute. The core of my objection to Hursthouse’s account of right action is that it obscures what is at issue. What is at issue is the normative condition of possibility for being virtuous. The normative condition of possibility for being virtuous depends on the existence of objective criteria for right action. An appeal to the actions of the virtuous agent achieves nothing in this regard because such an appeal either falls into the basic mistake of agent-basing or it becomes redundant.
The second objection to my critique of Hursthouse is that the appeal to the virtuous agent is not redundant because it is of some epistemic value when trying to discover what actions are right. I find this objection unpersuasive for the following reasons. First, Hursthouse’s (first) account of right action is a conceptual claim, not a practical claim. Her move from premise (1) to premise (2) in her account of right action is a response to the question, “What is a virtuous agent?” If her concern at this stage of her account was a practical concern, then it seems to me that the appropriate question to ask is “Why, under this description, is this action right?” or “Why does the virtuous agent perform this action?” This question is a demand for education, for an account of the point and purpose of right action, for an account of the nature of the good to such an extent as we can muster as well as an account of the actions that are constitutive of that good. Instead of asking this practical question, Hursthouse asks the conceptual question, What is a virtuous agent?

Admittedly, Hursthouse’s account of v-rules is a reply to practical concern. The objection raised the possibility that it is of some epistemic value to know how the virtuous man would act without possessing the corresponding knowledge of why he or she would act that way. The claim that the virtuous agent acts as our criterion for right action raises the following possibility. A asks, “What would a virtuous person do in this situation?” B responds, “A virtuous person would do X in this situation.” A asks, “Why would a virtuous person do X?” B replies, “I have no idea why, but when I ask myself ‘What would Jesus do?’ I recognize that he would do X. For example, I recognize that Jesus would not drive an SUV.” However, of what epistemic value are the conclusions concerning what Jesus would do to an agent who
does not have a firm grasp of why Jesus would act as he would act? After all, Agent A may misinterpret the situation in which he finds himself, might misperceive the options available to him, might misunderstand the good or goods involved and their relevant value, or simply have a corrupted understanding of what a particular virtue demands. Christine Swanton objects to Hursthouse’s “‘qualified agent’ account of right action” along similar lines, pointing not to the rational limits of the average agent, but to the rational limits of the virtuous agent. No virtuous agent is omniscient. As a result, there are many things about which a virtuous agent might be ignorant. When faced with ignorance that cannot be overcome, it seems odd, as Hursthouse herself admits, to appeal for advice to someone who is ignorant of the ends and issues involved in a situation.

Therefore, there are two primary reasons why I think Hursthouse’s account of right action is of little epistemic value in the absence of an account of the good. I should point out that Hursthouse, in Part III of On Virtue Ethics, does try to provide a naturalistic account of the good. However, she is quite explicit that the account of the good is designed to provide an account of the relationship between the good and the virtues – that is, the naturalistic account of the good provided in Part III is her account of teleology. She explicitly denies that the account of Part III is to be interpreted as a good-based account of right action. She says, “according to the form of ethical naturalism I am defending, what is particularly evaluated are character traits, not directly, actions.”


108 Hursthouse, On Virtue Ethics, 227.
to agent-basing, which as I have been using the term is concerned with an account of right action.

My conclusion is that the normative condition of possibility for being virtuous requires a good-based, not an agent-based, account of the existence of objectively appropriate reasons for right action. Before moving on to a discussion of the psychological conditions of possibility for being virtuous, let me mention some further implications of what I have said for normative theory.

2.5 The Structure of Normative Theory

I have argued against the claim that agent-based virtue ethics provided a coherent account of right action. In Shelly Kagan’s language\textsuperscript{109}, I was arguing against the claim that agent-basing provides a foundational device. We must abandon the claim that what makes an action right is that it is performed by an agent with the morally relevant internal states. However, it may still be the case that in determining the status of an action, by which we must understand that status as being praiseworthy, blameworthy, or neither – it may still be the case that the status of the action in this sense is directly determined by the virtues of the agent. For example, suppose a courageous man is faced with a dangerous situation in which he does not risk his life for the sake of whatever good is at stake. In this case, we might be able to determine that his action was praiseworthy (or at least not blameworthy), given the circumstances, since we are confident that if he did not act, then he must have had good reasons for not acting. But again, his failure to act in the face of danger is not

right because he is courageous. What makes his action right is whatever reasons he can cite for his failure to act in the face of danger – such as his true belief that he would surely fail. Nonetheless, given that the agent is courageous, it appears we could reasonably claim that his virtue is the only directly relevant factor in our determining if his actions were worthy of praise or blame.

However, this appears to be an epistemological claim of the following sort: since we are not virtuous, we are not adequate judges of the reasons for action that the virtuous agent perceived and assessed before making his choice not to act. Since we know that the virtuous agent is a good judge of those reasons, when we, the non-virtuous, make a determination as to whether or not his actions were worthy of praise or blame, we are appealing directly to his virtue as the only relevant factor.\textsuperscript{110} However, it is the only relevant factor for us. If the actions of the courageous individual were being assessed by other courageous individuals, then the fact that the individual was known to be courageous would be irrelevant. These courageous individuals would consider the circumstances under which our courageous agent made his or her decision and try to assess the reasons for not acting in the face of danger in much the same way that the initial decision was made.

In this context, consider cases in which we either 1) reserve judgment or 2) judge in favor of an agent by appealing to his character (or our assessment of his character). In these cases we say things like, ‘He is a good person, who has in the past performed many acts of courage. He must have had his reasons for not acting in these circumstances.’ We may ask him to explain what those reasons were, but

\textsuperscript{110} This point is related to point about fidelity and piety in 2.2.1.
ultimately, if we were not there, we may have to fall back on our assessment of his
character in order to determine whether his actions are blameworthy or not.

I am arguing that agent-based virtue theory may provide a plausible explanation of our assessments of actions as praiseworthy under conditions in which the agent is understood to be virtuous and those making the determination are not virtuous. To again put this point in Shelly Kagan’s language, agent-basing does not provide a coherent foundational device, but it may provide an account of the role virtue plays as an evaluative focal point.

2.6 Conclusion

Being virtuous involves the possession of character traits that provide an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability to act reliably for, or because of, those reasons. In order for any particular character trait to provide an agent with a capacity to consistently recognize objectively appropriate reasons for action of a certain kind, objectively appropriate reasons for action of an appropriate kind must exist. Therefore, an account of how it is possible for a human being to be virtuous presupposes the existence of objectively appropriate reasons for action. I have argued that agent-based virtue theories are unable to account for the normative condition of possibility for being virtuous.

An effort to provide an account of the normative condition of possibility for being virtuous is in part a quest for an account of the minimal amount of knowledge required to allow for one of the psychological conditions of possibility, namely, the capacity of human beings consistently to recognize objectively appropriate reasons
for action despite all their rational limitations. Whatever else this knowledge might be, it must involve some minimal conception of what is good or best. Some minimal conception of the good or the best is required in order to begin to speak coherently of appropriate reasons for action. Without a coherent way to talk about appropriate reasons for action, there is no coherent way to talk about virtue, being virtuous or right action. Nor is there any way to coherently speak about the psychological conditions of possibility for being virtuous.

My primary concern in Chapters Three and Four will be with the psychological conditions of possibility for being virtuous, accounts of which are necessary conditions for a plausible virtue ethics. To give these accounts is to meet Anscombe’s Demand.
CHAPTER 3

THE TEMPORALLY EXTENDED PLANNING THEORY OF HUMAN AGENCY

3.1 Introduction

In the last chapter I argued that Slote, Hursthouse, and others fail to account for the normative condition of possibility for being virtuous by appeal to a theory of agent-basing. I suggested toward the end of that argument that the existence of objectively appropriate reasons for action requires an appeal to a good-based account. On the good-based view, the evaluation of human actions is derived from and dependent on what we have to say about the good such that the primary source of objectively appropriate reasons for action is something beyond the mere traits of an agent.

Of course, I do not deny that a virtuous agent’s traits of character are the source of her having objectively appropriate reasons for action. What I am denying is that the source of the reason’s objective appropriateness is the agent’s traits.111 The argument of Chapter 2 was a denial of the latter normative claim, not the former psychological claim. In fact, I aim to affirm the psychological claim. I want to argue

111 Robert Audi draws a distinction between a reason state and a reason proper. A reason state is a psychological entity that can be a causal factor in action. A reason proper is an abstract entity that cannot be a causal factor in action. However, an action can be made intelligible (or be explained) in light of reference to a reason proper. My claim, then, is that the possession of a particular trait of character cannot be a reason proper, though it may contribute to the development of a reason state. See the opening chapter of Robert Audi, Action, Intention, and Reason (Ithaca: Cornell UP, 1993).
that it is an agent’s traits of character that allow him or her to correctly respond to reasons for actions which are objectively appropriate independently of his or her character. So while there must exist objectively appropriate reasons for action in order for an account of virtue ethics to be vindicated, there must also exist psychological traits that allow human beings to react correctly to those reasons. It is important to address the psychological claim separately from the normative claim because one may argue that even if objectively appropriate reasons for action exist, and even if human beings have the requisite knowledge concerning the good which would allow them to identify these reasons, it may still be the case that human beings are incapable of consistently recognize these objectively appropriate reasons for action and unable to reliably act for, or because of, those reasons.

Recall that I began with the following characterization of being virtuous:

Being virtuous involves the possession of character traits that provide an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability reliably to act for, or because of, those reasons. This understanding of being virtuous presupposes that despite all the limitations with which human beings are saddled, they nonetheless possess the capacity to consistently recognize objectively appropriate reasons for actions and possess the ability to reliably act for, or because of, these objectively appropriate reasons. I am calling the truth of these presuppositions the psychological conditions of possibility for being virtuous. My task in this Chapter is to attempt to provide the resources for an account of the psychological conditions of possibility for being virtuous. If I can provide an account of how it is possible for human beings to have these capacities and abilities, then I
will have provided an aid to the virtue theorist who wants to defend virtue ethics as a coherent moral theory.

I will argue that Michael Bratman’s temporally extended planning theory of human agency provides resources for accounting for the functional roles traits of character must play in order for being virtuous to be possible. The majority of this chapter is devoted to an articulation of the temporally extended planning theory of human agency and the resources it provides for accounting for the psychological conditions of possibility for being virtuous. As we will see, Bratman’s account of human agency stems from the basic premise that human beings are planning agents. We deliberate in advance of the moment of action in order to overcome serious limitations in our rational abilities. Bratman argues that most other accounts of human agency fail to explain this feature of our rationality and its most prominent feature, our present commitment to future actions as embodied in our intentions.\(^{112}\)

---

\(^{112}\) Much of Bratman’s work, especially his earlier work, is centrally concerned with an analysis of the relation existing between intentions and intentional action. Since my concern here is developing an account of traits of character, I will not be concentrating on the connection between intentions and intentional action. However, it will be impossible to avoid the issue entirely. After all, as Richard B. Brandt has pointed out, “traits of character…are concerned with intentional action” (Richard B. Brandt, “Traits of Character: A Conceptual Analysis” American Philosophical Quarterly, Vol. 7, No 1 (January 1970), 27). Therefore, we need to be able to explain some intentional actions, specifically virtuous actions, in terms of traits of character. In the account that follows virtuous action is explain by reference to intentions. However, I am not convinced that we need to follow Bratman in accepting the view that all intentional actions are to be explained by reference to intentions. See Ann Bumpus, “Actors without intentions: The Double Phenomena View” Philosophical Studies, 103 (2001): 177 – 199.
3.2 Intentions, Plans, and Commitment

In “Simple Intention,”[^113] Bratman attempts to provide an account of holding a non-explicitly conditioned future-directed intention. His emphasis is on the nature of the special form of commitment characteristic of this kind of intention, which he calls a simple intention. Bratman understands simple intentions as involving a kind of commitment that results in an agent “settling on” a particular course of action in the future.^[114]

According to Bratman, all intentions are recognized as conditioned in some way, either explicitly in the content of the intention or implicitly in the cognitive background of the agent. On the one hand, if any condition is part of the content of an intention, then the intention is conditional and one’s commitment to the action is explicitly conditioned as well. For example, my intention ‘to go to campus to work on my dissertation if the weather is too hot to allow working on the porch at home to be enjoyable’ is explicitly conditioned. I intend to go to campus only if it is too hot to work at home. By forming this intention, I am committed to going to campus, but only if it is too hot. It is not the case that I have now settled on going to campus. My future action is still an open question as far as I am concerned. Since my commitment to going to campus is explicitly conditioned, then the nature of my commitment cannot be characterized as my having settled on that course of action.


My commitment is explicitly contingent upon certain antecedent conditions being met.

If those conditions are not a part of the content of my intention, but only part of my cognitive background, then the intention is a simple intention in Bratman’s sense, and I can be said to have settled on that course of action. For example, suppose that during the month of July the heat in South Bend, IN is oppressive. Provided I know that the weather in South Bend is awful, I may form the intention ‘to go to campus to work on my dissertation.’ In such a situation, my intention to go to campus is not explicitly conditioned. My cognitive background contains the proposition ‘the weather during July is awful.’ Therefore, by forming the intention ‘to go to campus to work on my dissertation’ I am committed to going to campus. I have settled now on this course of action. Of course, my circumstances might change and as a result I may reconsider my intentions concerning a trip to campus. Nonetheless, I am not currently suspending my deliberation concerning my future action in order to determine what set of possible antecedent conditions are met. Rather, I regard my mind as made up.

Bratman wants an account of the transition from a conditional intention to a simple intention; that is, he wants an account of the conditions cited in one’s conditional intention being “pushed” into the background such that a conditioned intention becomes a simple intention to which one is committed in a manner one is not committed to a conditioned intention. What accounts for my mind being made up in the latter case, but not in the former?
Bratman entertains and rejects what he calls the cognitive fit view of simple intentions. The cognitive fit view of simple intentions “suggests that simply intending is just… intending under those conditions one expects.” On this view, the transition from conditional intention to simple intention is explained by the fact that the agent comes to believe that the qualifying condition of his conditional intention has been or will be met, and as a result the condition is pushed into the background to operate as a practical assumption. This sort of view is suggested by the description of my simple intention to go to campus due to persistent bad weather during July. Since I expect the weather to be awful, I settle on going to campus to avoid the heat.

Bratman rejects the cognitive fit view. His rejection of the cognitive fit view hinges on an appeal to an agent’s plans and the commitments entailed by those plans. Bratman utilizes the following counterexample to the cognitive fit view while drawing out the centrality of plans and commitment to his account of non-explicitly conditioned future-directed intentions.

Jill is a prisoner who has today a plan to dig a tunnel from her cell a little bit at a time for the next thirty days during dinnertime. She also has today the belief that she will be caught prior to the completion of her thirty day plan. She expects to be caught before fully executing her plan. So, Jill has today at least the following


116 Bratman’s appeal to planning in this argument is the seminal move in the development of the temporally extended planning theory of human agency. This is the first use of the concept of a plan that I have found in Bratman’s corpus.
conditional intentions: to eat in her cell on day thirty (if she is caught) and to finish her tunnel on day thirty (if she is not caught). However, since today she fully expects to be eating in her cell on day thirty, rather than finishing her tunnel on day thirty, according to the cognitive fit view, Jill has today the simple intention to eat in her cell on day thirty. Jill’s mind is now made up with respect to her future dinnertime activities; she now has settled on eating in her cell on day thirty.

But recall that according to Jill’s plan, she will be digging on day thirty. Her plan commits her now to digging on day thirty; that is to say, in virtue of her plan, she has now settled on digging on day thirty. Therefore, we cannot say now that she has a simple intention to eat in her cell on day thirty because we cannot say that she is appropriately committed to eating in her cell on day thirty. We can say that Jill has a conditional intention to eat in her cell on day thirty if she is caught prior to day thirty. We can also say she fully expects to be caught prior to day thirty. Nevertheless, we cannot say that she now simply intends to eat in her cell on day thirty because we cannot say of her now that she is appropriately committed to so acting then. The existence of Jill’s plan prevents the application of the cognitive fit view of simple intentions. As an alternative to the cognitive fit view, Bratman proposes that an account of simple intention must appeal to the commitments characteristic of one’s plans.

One might reply to this counterexample the way Robert Audi has.117 Audi’s strategy is to emphasize the distinction between hoping and intending. On Audi’s account, since Jill does not now expect to be digging on day thirty, we are only

---

117 See, for example, Audi, “Intention, Cognitive Commitment, and Planning,” 363.
entitled to say that she *hopes* to dig on day thirty or, alternatively, that she intends *to try* to dig on day thirty. We cannot say that she now simply intends to dig on day thirty. The cognitive fit view entails that an agent have a minimal level of belief that he or she will actually perform the action described in the intention. Therefore, if one believes she will not perform the action, we cannot claim that she intends the action. On the other hand, Audi asserts that hoping to perform an action is compatible with the belief that one will not perform the action.

However, consider Audi’s example of hoping. Audi is hoping to go to a concert, but believes he probably will not go because he cannot get tickets. In this situation we cannot claim that Audi intends to go to the concert. If, however, he were to receive tickets as a gift, we may immediately claim that Audi intends to go to the concert. The transition occurs because Audi now has the belief that he will go to the concert since he now has the tickets (and continues to have a desire to attend).

Notice the difference between Audi’s hope and Jill’s intention. Audi has no plan requiring that he secure tickets to the concert (despite the probability of failure) whereas Jill does have a plan requiring that she dig on day thirty (despite the probability of failure). One might suppose that the difference between Jill intending to dig and Audi hoping to go to the concert is that Jill has a plan that requires digging whereas Audi has no plan at all with respect to securing tickets.

Even so, one might reply that the existence of Jill’s plan only allows us to claim that Jill intends to try to dig on day thirty, not that she intends to dig on day thirty. Audi, on the other hand, prior to receiving the gift of the tickets, does not even


88
intend to try to go to the concert, he merely hopes to go. Therefore, we might characterize the difference between hoping and intending to try has dependent on a plan, but we cannot conclude that the difference between hoping and simply intending is dependent on a plan. It remains odd to claim that an agent simply intends to perform an action required by a plan while believing that he or she will not perform it.

Audi’s attempt to salvage the cognitive fit view with an appeal to the distinction between hoping, intending to try, and simply intending, fails for the following reasons. According to Audi, to plan to A entails that one intends to A. However, to intend to A entails that one believe that one will A. Therefore, to plan to A entails that one believe that one will A. It follows that if one does not believe that one will A, then one cannot be said to plan to A.

But this claim is objectionable. One’s beliefs concerning success or failure have no necessary connection to the plans one is capable of forming, not even to the rational plans one is capable of forming. It is possible, and often preferable, desirable, and rational to form plans despite the fact that one believes one will fail. In fact, Jill’s case demonstrates not only the possibility, but the appropriateness of planning to perform an action despite one’s belief that one will fail to perform it. If Jill had realized that because she did not believe she would dig on day thirty, she could not (rationally) intend to dig on day thirty, and because she could not (rationally) intend to dig on day thirty, she could not (rationally) plan to dig on day thirty, then she would have realized that she could not (rationally) plan to dig herself.

out of her cell. But to claim that she could not (rationally) plan to dig herself out of her cell is absurd. Her plan may very well turn out to be futile, but she can, and should, form such a plan. If she were not to have formed such a plan, she would have instead resigned herself to captivity.

The role of a belief condition in an account of intention remains controversial. It does seem that one cannot intend to perform an action that one believes is impossible, such as walking unaided on the ceiling. It also may seem that one cannot intend an action or to bring about a state of affairs, such as good weather, over which one believes he or she has no control or influence. So, while the cognitive fit view does not capture the proper cognitive constraints on intending, there is still significant room available for cognitive constraints on intending. An important question is whether the appropriate kinds of cognitive constraints must be constraints according to one’s beliefs.

It is often argued that “in practical reasoning and action we seek to realize our intentions and satisfy our desires ‘in light of what we believe.’” Reasonable belief is usually taken to have the following features: 1) reasonable belief is context independent; 2) reasonable belief is shaped primarily by a) evidence for what is believed and b) concern for the truth of what is believed; 3) belief is not normally

---


subject to direct voluntary control; and 4) an agent’s beliefs are subject to integration.\(^\text{122}\)

In “Practical Reasoning and Acceptance in a Context,” Bratman argues that each of these features of belief do not necessarily belong to the cognitive attitude operating in practical reasoning. The reason for this dissimilarity, as we saw in the case of Jill, is that various kinds of practical pressures require an agent to take for granted or assume in the background of his or her practical reasoning propositions that he or she does not believe. These practical pressures are context relative, and for the purposes of his or her practical reasoning, an agent will accept certain propositions based on these context-relative pressures even though he or she may not believe the propositions in context independent terms. For example, independent of her plan, Jill does not now believe she will be digging on day thirty. However, within the context of her plan, she now accepts that she will be digging in her cell on day thirty. The practical pressures cited by Bratman include pressures to simplify one’s reasoning, asymmetries associated with cost of errors in different contexts, needs for social cooperation, and the maintenance of certain personal relationships. Bratman offers the following example of operation of acceptance in a context,

\[\text{A soldier in a war zone has his doubts that he will make it through the day and expresses these doubts in a letter he writes in the morning. Nevertheless, after writing his letter, he proceeds to make plans for his daily tour of the battlefield; and in so doing, he takes it for granted that he will be around to execute those plans. After all, how else could he plan for the day? Since he needs to make such an assumption in order to get his planning}\]

\(^{122}\) These four features of belief are taken from Bratman, “Practical Reasoning and Acceptance in a context,” 27.
off the ground, such acceptance may be reasonable even in the face of his doubts.\textsuperscript{123}

Examples such as this show, according to Bratman, that effective and rational practical reasoning is not necessarily limited by what an agent believes because what one accepts/takes for granted (a) can reasonably vary, … across contexts; (b) can be influenced by practical considerations that are not themselves evidence for the truth of what is accepted; (c) can be subject to our direct voluntary control; and (d) is not subject to the same ideal of [integration] across contexts.\textsuperscript{124}

Therefore, the cognitive attitude often operating in practical reason is not necessarily belief but rather often acceptance. We can conclude then that the reason the cognitive fit view fails as an account of simple intention is because it fails to acknowledge that the cognitive attitude operating in one’s plans is often acceptance.

It is true that the default background of deliberation is the set of an agent’s beliefs. However, this background can be altered, sometimes radically, by the acceptance of certain propositions due to practical pressures of various sorts. The resulting context-relative adjusted cognitive background is the result of positing for acceptance certain propositions to serve as part of the deliberative background, while also bracketing the influence of certain beliefs on one’s deliberation. Therefore, I propose that appropriate cognitive constraints on intending are that the agent comes to accept in the context of his or her plan that the qualifying condition of his intention

\begin{flushright}
\textsuperscript{124} Bratman, “Practical Reasoning and Acceptance in a Context,” 27.
\end{flushright}
has been or will be met, that the agent believe the action is possible, and that the agent believe that he or she has at least some potential, though perhaps indirect, influence on the intended action or state of affairs.

Recall that the issue of cognitive constraints on intending arose out of an attempt to account for the proper kind of commitment characteristic of an intention. The rejection of the cognitive fit view depended on an appeal to the plans of the agent and the commitment embedded in his or her plans. However, the appeal to an agent’s plans did not rule out cognitive constraints on intending, but rather helped draw out the practical pressures operating in an agent’s reasoning with respect to his or her plans. An examination of these pressures revealed that the commitment embedded in one’s plans does involve an appropriate cognitive attitude, though not necessarily a belief.

While the role of a cognitive constraint on intending may still remain controversial, what does not appear overly controversial is that commitment plays an important role in having a simple intention. Consider again Jill’s situation. Jill is committed to escaping captivity. It is this commitment to escaping, not her beliefs concerning her likelihood of success or failure, which allows her to plan her escape and forces her to accept certain propositions, thereby having the simple intention to dig on day thirty.

3.3 Commitment

The concept of commitment and the embodiment of commitments in intentions and plans will form the trunk of my account of traits of character. In nearly every instance, my account of how the psychological conditions of possibility for
being virtuous are met will make an appeal, either directly or indirectly, to the notion of commitment. Before moving on it is appropriate to say a bit more about commitment, both in general and specifically on Bratman’s account of human agency.

The literature on commitment is growing, especially in relation to the work by Charles Taylor, Bernard Williams, Michael Stocker, and others concerning the relationship between, on the one hand, one’s commitments to values, norms, projects and persons, and on the other hand, one’s self-understanding, identity, integrity, and self-worth. Marcel Lieberman has offered the following useful characterization of the central features of commitment:125

1. its stability over time and its capacity to be revised and reconsidered;126
2. its action-guiding force;127
3. its relation to self-understanding and identity.128

125 Lieberman, Commitment, Value and Moral Realism, 5.
In addition to Lieberman’s initial characterization of commitment we should also recognize:

4. its practical necessity.129

To say that a commitment is stable over time is to recognize that momentary feelings, thoughts, or actions are insufficient to characterize an agent’s attitude as that of commitment. In addition, commitment implies that the object of commitment will have a significant controlling influence over an agent, both on his or her deliberation and on his or her actions, for an extended and uninterrupted amount of time. One who is committed does not succumb to temporary preference changes, temporal discounting, temptation, or social pressure, and in general is able to remain temporally consistent with respect to his or her present and future plans, projects and goals. However, temporal consistency as exhibited in the stability characteristic of commitment does not entail that one’s commitments, at least in the sense that I am using the term, cannot be revised. There is room for confusion here, so a distinction is in order. Most commitments are not like promises; most commitments do not generate specific obligations either to oneself or to another.130 Instead, an agent is capable of revising his or her commitments. However, when revision of one’s commitments does occur, the concept of commitment requires that the revision be the


130 Schaub, Kupperman, and Lieberman all emphasize this point. Each, however, has provided a different vocabulary for discussing the distinction. None I think is appropriate, and so I have chosen not to adopt any of their uses. By contrast, as van Hooft has pointed out, Robert Solomon has advocated the view that commitment is a form of promise. See Solomon’s Love: Emotion, Myth and Metaphor (New York: Anchor Press Doubleday, 1981), 224.
result of either reflective deliberation or a personal (moral) crisis accompanied by powerful affect.\textsuperscript{131} The notion of commitment prohibits the revision of commitment on the basis of changes in one’s present dominant desires and rarely allows revision on the basis of mistaken beliefs or unexpected circumstances.

Commitments are action-guiding in the sense that to be committed to something is to give the object of that commitment a prominent place in one’s deliberations, or alternatively, to allow the object of that commitment to shape and direct one’s deliberation. An agent regards the object of her commitment to be important to such an extent that she accords “deliberative priority to projects consistent with it.”\textsuperscript{132} By being committed to something, one regards certain parameters of practical reasoning as set, particularly ends and goals, but also acceptable means. By orientating and constraining the direction of practical reasoning, an agent’s commitments are action-guiding. In addition, typically when we speak of an agent being committed to something, we understand that commitment to be a positive commitment. We understand a committed agent to be not only willing to act positively when required, but also to not attempt to avoid situations that might potentially require action by the agent.\textsuperscript{133} Some types of commitment might

\textsuperscript{131} See Haste, “Moral Responsibility and Moral Commitment” for an account of three different avenues toward the formation of commitments.

\textsuperscript{132} van Hooft, “Obligation, Character, and Commitment,” 361. Deliberative priority is a concept deployed by Bernard Williams in \textit{Ethics and the Limits of Philosophy}.

\textsuperscript{133} This qualification of the action-guiding character of commitment is necessary to avoid what I think is a misguided way of thinking about commitment. Sarah Stroud, in “Moral Commitment and Moral Theory” has argued in favor of the counterfactual-condition model of moral commitment. She insists on the counterfactual-condition model because a material-conditional model would allow the material conditional to be true, and therefore the agent committed, whenever the antecedent of the conditional was false. So, for example, “for all x, if I meet x I treat x courteously” would always be satisfied provided I never met anyone. It follows then, that one can be committed to courtesy so long
also require actively seeking out opportunities to act in ways consistent with the commitment.

   Commitment also produces practical necessity in the following sense. “It seems to be a characteristic of commitment that the motivation to act is not so much a sense of obligation as a sense of compulsion.”\textsuperscript{134} When an agent is committed to something, and action is required to conserve or maintain the object of the commitment, the agent experiences the required action as something he or she must do.\textsuperscript{135} What is more, the object of commitment plays a central role in the practical deliberations of the committed agent. The object of the commitment is taken to be important and is given deliberative priority such that it is not an appeal to duty or obligation that explains the actions of the agent, but rather the circumstances surrounding the object of the commitment. We are reluctant to say that someone is truly committed to something if he or she is either not motivated to act in ways consistent with the commitment or is only motivated out of a sense of duty.\textsuperscript{136}

   Finally, our commitments constitute our identity, allowing for self-understanding, providing for integrity, and establishing standards of self-worth. Stan

---

\textsuperscript{134} Haste, “Moral Responsibility and Moral Commitment,” 319.

\textsuperscript{135} van Hooft in “Obligation, Character, and Commitment” has argued that Bernard Williams’ account of practical necessity (obligation) in terms of importance and deliberative priority must be supplemented with commitment.

\textsuperscript{136} Michael Stocker has emphasized how acting from a sense of duty is often antithetical to the commitment we expect in, for example, personal relationships such as friendship. See Michael Stocker, “The Schizophrenia of Modern Moral Theories” in \textit{Virtue Ethics}, ed. Roger Crisp and Michael Slote (New York: Oxford UP, 1997), 66-78.
van Hooft has defined commitment as “an attitude which gives to its object a positive and practical importance which involves that object in one’s own integrity. One invests oneself in what one is committed to.”\textsuperscript{137} Our sense of self, our ability to make sense of our lives as a unified whole, our ability to recognize continuity and temporal extension within our lives are each linked in important ways to our commitments to plans, projects, practices, institutions, traditions, and, perhaps most centrally, significant others such as family, friends, and colleagues. Joel Kupperman, as well as many others, has emphasized the central importance of a narrative unity to a meaningful human life.\textsuperscript{138} One’s commitments to objects of importance provide the structural elements for a narrative unity of a human life. It is by living one’s life out within a narrative structure that one is able to reach self-understanding, that one is able to develop the ability to deliberate in ways that will maintain one’s integrity, and that one is able to assess one’s own success or failure in living up to one’s commitments.

Given these four characteristics of commitment, one might ask what accounts for them. One possible response is to understand commitment of the sort described here in relation to simple intentions and plans.\textsuperscript{139} As we saw above in the case of the prisoner Jill, one’s plans involve a characteristic kind of commitment. According to

\textsuperscript{137} van Hooft, “Commitment and the Bond of Love,” 456.
\textsuperscript{139} For a good discussion of this strategy, see Lieberman, Commitment, Value and Moral Realism, Chapter 3. Lieberman rejects this strategy for reasons I will discuss below in Chapter 5.
Bratman, this commitment involves two dimensions: reasoning-centered commitment and volitional commitment.\textsuperscript{140}

Reasoning-centered commitment is characterized by inertia and consistency.\textsuperscript{141} Inertia of intention consists in a disposition on the part of the agent to retain an intention without reconsideration except when faced with a “problematic situation.” A problematic situation is one in which the beliefs or acceptances upon which one formed an intention turn out to be mistaken to such an extent that they can no longer operate within the agent’s practical reasoning. Alternatively, a problematic situation can arise when the circumstances in which one finds oneself make the execution of one’s intentions impossible or highly imprudent. Since Jill’s plan is to escape from captivity, she is disposed to retain her intention as embodied in her plans. Her commitment to escape is manifested by her failure under her normal conditions to reconsider her plan. Once on a particular course of action, she is disposed to continue on that path unless something extraordinary occurs to cause her to reconsider her plan. Inertia lends stability to one’s intentions. In the primary sense, stability concerns a disposition not to reconsider. In a secondary sense, stability concerns a disposition intentionally to refrain from reconsidering.

The consistency component of reasoning-centered commitment involves a disposition to reason from prior intentions to further intentions while maintaining consistency between the prior and the further intentions by way of constraining one’s


\textsuperscript{141} See Bratman, \textit{Intentions, Plans and Practical Rationality}, 16-20.
deliberation and the formation of further intentions. For example, once Jill has settled on escaping from captivity, her intention to escape limits the range of her deliberation to actions consistent with the execution of her plan. Eating in her cell on day thirty is inconsistent with her plan of escaping, and therefore her commitment to escaping will not allow her to form the intention to eat in her cell on day thirty.

Volitional commitment results from a conduct-controlling pro-attitude. Volitional commitment results from a conduct-controlling pro-attitude. A conduct-controlling pro-attitude cannot be reduced to a desire since desires are merely potential conduct-influencing pro-attitudes. A conduct-influencing pro-attitude may affect one’s conduct at a particular time, but such a pro-attitude must be ‘weighed’ against other pro-attitudes one may have at the time. This is to say, one may have a conduct-influencing pro-attitude toward X, but in the end not perform X. On the other hand, a conduct-controlling pro-attitude can affect one’s actions without being weighed against other conflicting pro-attitudes. This is to say, in the normal course of events, if one has a conduct controlling pro-attitude to X, one will X. A simple intention, then, need not, and often should not, be weighed against conflicting desires. Typically a rational agent will not deliberate in such a way that her simple intentions are reconsidered. When one regards oneself as having settled on a course of action, one’s conduct is typically controlled by one’s intentions without further deliberation.

These two kinds of commitment, reasoning-centered and volitional, are closely related and in the end tied to action. While the reasoning-centered

commitment is concerned directly with dispositions concerning our reasoning, such reasoning allows us to gain volitional commitments.

Intentions, then, are stable insofar as they resist reconsideration due the inertia characteristic of reasoning-centered commitment, are action-guiding insofar as intentions provide inputs into practical reasoning due to the consistency characteristic of reason-centered commitment, are sources of practical necessity due to volitional commitment, and contribute to the unity and integrity of a human life insofar as intentions are constitutive elements of one’s plans and projects.

We are now in a position to begin to develop an account of the conditions of possibility for being virtuous. This account will be provided in terms of the functional roles played by the virtues. These functional roles, I will argue, can be satisfied in part by an appeal to the commitment characteristic of unconditional future-directed intentions as understood in Bratman’s temporally extended planning theory of human agency.

3.4 The Temporally Extended Planning Theory of Human Agency

Human beings are planning agents. We are under various kinds of pressures to be planning agents. One central pressure stems from our bounded rationality. Bounded rationality is characterized by the fact that our decisions usually concern discreet, specific matters and do not range over our entire lives, our decision making process does not include exhaustive probabilistic characterizations.

143 See Bratman, “Taking Plans Seriously” and Intentions, Plans, and Practical Reasoning.

of all the possible alternative futures, and our decision making process usually involves a type of selective attention in which certain facts about our selves and our lives are ignored. H. Simon calls a theory derived from these observations the behavioral model of bounded rationality. Bounded rationality allows an agent to give his or her attention to certain extremely important variables in a situation while ignoring the vast number of possible influences on his or her decisions. Our need for the ability to form and execute plans is the result of the limitations of our abilities to deliberate and certain limitations in our knowledge resulting from our bounded rationality. Since bounded rational agency of the kind humans have requires deliberation, and deliberation under the conditions of bounded rationality only at the time of action would be under such time constraints that its value would be minimal, humans need a way for deliberation and rational reflection to influence action beyond the present. The formation of simple intentions and plans fulfills this need. Settling now on a plan enables present deliberation and reflection to influence an agent’s later conduct. If circumstances approximate what one expected them to be when one formed a plan, then, at the time of action, there is no need for deliberation as opposed to execution of one’s pre-existing plan.

In addition to compensating for our bounded rationality, plans help compensate for other features of human nature. As temporally extended and social agents, humans need plans to facilitate inter-personal coordination in order to achieve complex goals. In the face of equally desirable options, human beings have the psychological capacity to ‘pick’ one option or the other. If I know only your beliefs

and desires, it would be very difficult for me to anticipate what you are going to pick. If I know your intentions, the range of what is equally desirable may thereby be diminished and my ability to anticipate what you will do is increased. The increased ability to anticipate what you will do helps with inter-personal coordination.

Due to the pressures resulting from bounded rationality and our natures as social creatures, human beings are planning creatures. The central component of our natures as planning agents is the possession of certain mental states called intentions. Typically an intention is understood to be a desire-belief complex that is the result of practical deliberation. However, such a desire-belief model of intention is rejected by the temporally extended planning theory of human agency.

3.4.1 Rejecting the Desire-Belief Model of Human Agency

As Bratman developed his temporally extended planning theory of human agency, he develops increasingly pointed critiques of what he calls the desire-belief model, particularly the views of Davidson.\(^\text{146}\) Despite the apparent need for and nature of plans in human rationality, many scholars are resistant to the idea that there could be anything that plays the role of plans.\(^\text{147}\) Bratman characterizes this position as skepticism concerning future-directed intentions. Skepticism concerning future-directed intentions naturally leads to an account of intention which begin with notions of intentional action and acting with an intention rather than beginning with

\(^{146}\) For Davidson’s account of intentional action, see “Actions, Causes, and Reasons.” For his account of all-out evaluations see “How is Weakness of the Will Possible?” For his account of future-directed intentions see “Intending.” All three articles can be found in Davidson, Essays on Actions and Events. See Bratman, Intention, Plans, and Practical Reason, 6.

\(^{147}\) See Bratman, Intention, Plans, and Practical Reason, 3-9.
intending. This approach to the analysis of intention Bratman labels ‘the methodological priority of intention in action.’\textsuperscript{148}

Bratman argues that the priority given to intention in action leads naturally to a position he labels the desire-belief theory of intention in action. According to this view, we understand intentional action, and action done with an intention, in terms of an agent’s actions standing in appropriate casual or non-causal relationships with the agent’s desires (or pro-attitudes) and beliefs.\textsuperscript{149} Given the methodological priority of intention in action, the desire-belief theory of intention in action usually deals with the notion of intending and future-directed intentions by reducing future-directed intention to complexes of desire (or pro-attitude) and belief. On these reductive accounts, there is no special mental state or attitude of intending to act. The desire-belief model of intention in action and intention reduction constitutes the popular desire-belief model of human agency. There are two popular versions of reduction of an intention to a desire-belief pair. The temporally extended planning theory of human agency rejects both versions.

3.4.1.1 Reductive Conception of Intention: Version 1

The first reductive conception of intention seeks to argue that an intention is merely a strongest desiderative reason. On such a conception, we cannot coherently say that when one forms an intention to X, one has a further reason to opt for the


\textsuperscript{149} Anscombe is a non-causal advocate of the desire-belief theory of intention in action while Goldman and early Davidson are advocates of the causal version.
means to X. The reason for the means to X continues to be the desire to X. The formation of an intention to X is merely an output of practical reasoning, and therefore plays no role as an potential input. Bratman rejects this reductive conception of intention. As noted above, an intention involves both volitional and reasoning-centered commitment. Volitional commitment involves a conduct-controlling pro-attitude. Desire, on the other hand, is merely a potential conduct-influencing pro-attitude. Therefore, one cannot account for volitional commitment if one reduces an intention to a desiderative reason, no matter how strong.

Bratman also rejects this first reductive view because it fails to provide an adequate account of the reasoning-centered commitment of future-directed intentions. Reasoning-centered commitment involves inertia and consistency. Inertia of intention consists in a disposition to retain an intention without reconsideration. Consistency of intention involves a disposition to reason from prior intentions to further intentions while maintaining consistency between the prior and the further intentions by way of constraining my further intentions. One way to attempt to capture the reasoning-centered commitment of future-directed intentions on the desire-belief model is to reduce future-directed intentions to present predominant desire. However, Bratman argues, a predominant desire cannot guarantee reasoning-centered commitment. A predominate desire does not guarantee reasoning-centered commitment because it does not guarantee that I will not continue to deliberate about whether or not the object of my predominate desire is what I will pursue. As Bratman says, I do not see the issue as settled. There is nothing about having a predominate desire that ensures that I will have a disposition to not reconsider pursuing it.
Similarly, there is nothing about a predominate desire that ensures that I will have a disposition to reason about the means to achieving it. Nor is there anything about predominate desires that ensures that I will have a disposition to reason in such a way that further intentions are consistent with the prior predominate desire.

3.4.1.2 Reductive Conception of Intention: Version 2

Another reductionist strategy is to claim that to intend to X is to have a predominate desire to X and to believe that one will X (because of this desire to X). However, the addition of the belief condition does not help the desire-belief model because the addition of the belief condition does not guarantee that I regard the issue of what I should do as settled. Bratman offers the following counterexample to this reductionist thesis. Suppose I have a predominant desire to go to the library and, because I know my work habits, I also believe I will go. However, I also regard my desire (and so my belief) to be the result of an objectionable form of addiction to work, and so I may continue to deliberate about whether or not to go to the library (all the time believing that I will go). In this type of case the addition of the belief condition does not guarantee that an agent will regard the supposed intention as a settled matter.

Bratman argues that any version of the desire-belief model of human agency does not provide the resources to understand ourselves as planning agents. Any account of intention must recognize the fact that future-directed intentions are formed as parts of plans. As parts of plans, intentions are not only performed in the course of executing a plan, but also play an important role as inputs into further practical reasoning. Future-directed intentions play a role in modifying, specifying,
constraining, reconsidering, and retaining plans. The relation of intentions to plans and actions is complex, but the desire-belief model of intention simplifies this relationship, often rendering future-directed intentions nothing more than outputs of practical reasoning to be executed. Furthermore, the desire-belief model of intention has no room for distinctive attitudes or states of intending to act. Bratman will argue that we cannot make sense of ourselves as planning agents unless we recognize a distinctive mental state on a par with beliefs and desires.

3.4.2 Description of the Temporally Extended Planning Theory of Human Agency

Bratman’s basic assumption in the development of the temporally extended planning theory of human agency is a functionalist philosophy of mind.\textsuperscript{150} Functionalism is the view that what defines a mental state is the relation between certain inputs into mental activity and certain outputs. We understand a mental state S as being a particular mental state based on the role it plays in the activity of an agent. Our commonsense understandings of mental states require positing a network of underlying regularities in which these mental states are embedded. “These regularities connect these various states with each other, with associated psychological processes and activities, and with characteristic ‘inputs’ and ‘outputs’: perception and action.”\textsuperscript{151} So as to avoid certain mistakes involved in behaviorism, it is essential that the relevant regularities are not confined to dispositions to act in certain ways, but include dispositions to think and reason in characteristically human

\textsuperscript{150} See Chapter 1 and my discussion of a Wittgensteinian objection.

ways. As we saw, planning is a characteristic way in which human beings engage in bounded practical reasoning. The regularities of bounded practical reasoning need not be strict regularities, but rather “ceteris paribus laws” involving associated norms such as consistency. For example, when we have inconsistent beliefs we have some disposition to render them consistent. Associated with this disposition to modify our beliefs is a norm enjoining consistency in beliefs. Bratman believes there is a parallel concerning the consistency of intentions. The dispositions and associated norms are psychological features of rationally bounded human beings. The regularities that are of importance to an understanding of human agency are the regularities and dispositions that govern the bounded practical reasoning of planning agents who compensate for their limitations by forming volitional and reasoning-centered commitments. An account of human agency will require an account of the dispositions and associated norms that govern the reasoning of limited planning agents.

With these basic functionalist assumptions, Bratman proposes the planning theory of human agency. Central to the planning theory of human agency is the role intentions play as inputs into the practical reasoning of rationally bounded creatures such as human beings. Bratman considers two views of the way intentions operate as inputs into practical reason. The first is the intention-based reasons view and the second is the framework-reasons view.

The intention-based-reasons view holds that just as there are desire-belief reasons for action, so too there are intention-based-reasons for action. On this view, ____________

---

152 For Bratman’s account of functionalism see *Intention, Plans, and Practical Reason*, 9-10.
since intentions provide reasons for action just as belief-desire complexes do, we can say that intentions are directly relevant to practical rationality. But Bratman thinks this view is both too weak and too strong. It is too weak because it has the unacceptable consequence of allowing intentions to be one reason among many reasons in means-end reasoning. This consequence is unacceptable because by allowing an agent’s intention-based reasons to be weighed against other desire-belief reasons, we are allowing means-end reasoning to destabilize the intention. “To see my intention as providing just one reason among many is to fail to recognize the peremptoriness of reasoning-centered commitment.”153

The intention-based-reasons view is also too strong because it permits unacceptable forms of bootstrapping. Bratman insists that “we take seriously the idea that intentions are psychological elements that are distinct from desires and beliefs, and that can themselves generate reasons for and against further intentions.”154 However, Bratman thinks that taking this idea seriously, as well as its implication that intentions have independent reason giving force, can lead to illegitimate bootstrap rationality. The generation of the problem depends on the assumption that to act rationally is to act on that decision arrived at by ‘adding up’ all the reasons for an action that the agent has. On the first reductive account mentioned above this would mean determining which desiderative reason was strongest and acting accordingly. But since Bratman has allowed intentions to count as reason, they too must be included in the reckoning. If they are included, then it is possible that a prior


intention will ‘outweigh’ a strongest desiderative reason. This becomes problematic when the intention was the result of irrationality. Before forming the intention to X, one had to reckon the reasons for X-ing and the reasons for Y-ing. Suppose the reckoning deemed Y-ing rational, yet one still formed the intention to X. Now, when the intention to X is considered as a reason to opt for means to X or means to Y, it appears to be rational to opt for the means to X. Bratman argues that this is illegitimate bootstrapping. An action that is otherwise irrational has become rational.

Due to the problems inherent in the intention-based-reasons version of the planning theory, Bratman attempts to provide a more sophisticated view which we can call the framework-reasons view. Plans are, in Bratman’s sense, mental states. Specifically they are large intentions and so have the same properties as intentions: plans resist reconsideration (due to reason-centered commitment), control conduct (due to volitional commitment), and, perhaps most importantly, provide input into practical reasoning. Plans have two additional properties; they are partial and hierarchical. Plans must be partial in order to allow us to adapt to changing circumstances and must be hierarchical to help structure not only our days, but our lives as well. An agent’s volitional and reasoning-centered commitments and hierarchically arranged partial plans provide a framework within which bounded practical reasoning can occur.

---

155 See Bratman, *Intentions, Plans and Practical Reason*, 28-29. Strictly speaking, it is the “having” of a plan which is a mental state. For stylistic reasons, Bratman avoids the construction “having a plan.” Plans are “mental states involving an appropriate sort of commitment to action: I have a plan to A only if it is true of me that I plan to A.”


Since plans, like intentions, are to support coordination and systematically extend the influence of present deliberation to future conduct by providing a framework for bounded practical reasoning, plans are under a number of (rationally defeasible) constraints required if they are to successfully perform their functions.  

1) Internal consistency: plans should be internally consistent such that it is possible for an entire plan to be executed.  
2) Cognitive consistency: a plan should also be consistent with my beliefs and acceptances about how the world is such that it is possible for an entire plan to be executed if my beliefs and acceptances about the world are true.  
3) Positive means-end coherence: plans should also be constructed and formed in such a way that the relationship between one’s ends and one’s means is coherent. Means-end coherence also carries with it the demand to settle on some means to executing the intention or plan. “The demand for means-end coherence provides rational pressure for the addition of further intentions.”  

Bratman notes that not only do desires differ from intentions in that only intentions embody volitional and reasoning-centered commitment, but desires also are not subject to any of these (rationally defeasible) constraints. Desires need not be consistent with one another; they need not be consistent with our beliefs and acceptances, nor does having a desire to A require an agent to deliberate about means to fulfill that desire.

---

158 For each of the following, see Bratman, *Intention, Plans, and Practical Reason*, 31

159 Bratman, *Intention, Plans, and Practical Reason*, 33. This requirement of means-end coherence can be viewed as a reply to Nancy Schuaber’s objection to a material-conditional model of commitment.
Therefore, there are at least five features of intentions and plans that distinguish them as mental states from desires or desire-belief complexes:

1. Volitional commitment
2. Reasoning-centered commitment
3. Internal consistency
4. Cognitive consistency
5. Positive means-end coherence

Each of these features of intentions plays a central role in the operation of bounded practical reasoning.

Given that plans resist reconsideration (due to reasoning-centered commitment), plans pose problems for deliberation about further intentions and actions while simultaneously constraining the range of available options and directing an agent’s attention to acceptable options.\textsuperscript{160} Plans provide standards of relevance for options considered in deliberation in virtue of the demand for means-end coherence.\textsuperscript{161} In addition, the range of possible solutions to the problems posed by the demand for means-end coherence is constrained by the demand for cognitive consistency such that plans provide a filter of admissibility for options considered in deliberation.\textsuperscript{162} By generating relevant decision problems and more or less heavily restricting the range of possible solutions to those problems, plans and intentions make practical rationality possible for rationally bounded creatures. Without playing these functional roles in human practical reasoning, it is hard to see how limited creatures such as us could have any rationale for sifting through the limitless range of

\textsuperscript{160} Bratman, \textit{Intentions, Plans and Practical Reason}, 32.

\textsuperscript{161} Bratman, \textit{Intentions, Plans and Practical Reason}, 33.

\textsuperscript{162} Bratman, \textit{Intentions, Plans and Practical Reason}, 33.
options open to us at any given moment. Bratman summarizes the role of intentions and plans as follows:

Prior intentions and plans, then, provide a background framework against which the weighing of desire-belief reasons for and against various options is to take place. This framework helps focus deliberation: it helps determine which options are relevant and admissible. Prior intentions provide this background framework directly: the framework does not depend on the agent’s seeing her intentions as evidence for something else, or on the presence of some special desire – for example, to stick to her guns. Nevertheless, in playing these roles intentions do not provide reasons for action to be placed on the scale with desire-belief reasons in determining what to do [contra the intention-based-reasons view]. Their role is to help determine what options are to be considered in the process of weighing conflicting reasons for action rather than to provide reasons to be weighed in favor of one considered alternative over another. The reasons to be weighed in deliberation remain desire-belief reasons.\textsuperscript{163}

Intentions and plans provide framework reasons “whose role is to help determine the relevance and admissibility of options.”\textsuperscript{164} Not being forced to see intentions as providing reasons that are one reason among many is the framework-reasons view’s first improvement over the intention-based-reasons view.

The second improvement of the framework-reasons view over the intention-based-reasons view is that the framework-reasons view of the planning theory of human agency provides for a more complex view of practical reasons. On the

\textsuperscript{163} Bratman, \textit{Intention, Plans, and Practical Reasons}, 33-34

\textsuperscript{164} Bratman, \textit{Intentions, Plans and Practical Reason}, 34.
framework-reasons view, practical reasoning has (at least) two levels. On the first level, prior intentions and plans pose problems for deliberation and provide a filter on the range of possible solutions to those problems. On the second level, desire-belief reasons enter deliberation between relevant and admissible options. Two-tiered practical reasoning facilitates coordination by making behavior more predictable (when we want to be predictable) and prevents humans from being merely time-slice agents by recognizing the temporally extended nature of deliberation and choice.

The third improvement of the framework-reasons view of the planning theory of human agency is that it provides a solution to the problem of bootstrapping. It solves the problem by first distinguishing the perspective internal to the agent’s deliberation from the perspective external to the agent’s deliberation. The internal perspective includes the influence of the framework of the agent’s prior intentions and plans on his or her deliberation. The perspective external to the agent’s deliberation is a perspective from the agent’s desires and beliefs without the influence of his prior intentions and plans, but rather in light of the ultimate rationale for having intentions in the first place. On Bratman’s view, the ultimate rationale for having intentions in the first place is to facilitate human beings of bounded rationality in the satisfaction of (rational) desire. The solution to the bootstrapping problem is to


166 For a discussion of the centrality of acknowledging in an account of character the temporally extended nature of choice, see Joel Kupperman, *Character*.


168 Bratman, *Intentions, Plans and Practical Reason*, 43. Here we see Bratman’s endorsement of a desire-satisfaction account of foundational reasons for action. In Chapter 4 I will reject this account in favor of a good-based foundational account of reasons for action consistent with the normative account of possibility for being virtuous as discussed in Chapter 2.
recognize that while it might be rational for an agent to act on a certain desire-belief reason in light of his prior intentions, these intentions can be criticized from the perspective of the agent’s desires and beliefs and their relationship to the agent’s (rational) desire satisfaction. From a perspective unconstrained by the framework of the agent’s prior intentions and plans, ‘we’ can ask about any intentional action whether it was rational with respect to the (rational) desire satisfaction of the agent. It is possible for ‘us’ to determine that an option ruled inadmissible within the framework of the agent’s prior plans and intentions was, in fact, a superior option for the agent given his (rational) desires and beliefs. The assessment from the external perspective is only an assessment of the rationality of a particular action at a particular time, not an assessment that the agent should have performed the action that is determined to be more rational. Indeed, it may have been irrational for the agent to enter into the deliberation required to determine that, in fact, this alternative course of action was the most rational. Such deliberation may have required an irrational and time-consuming reassessment of all or many of the agent’s prior plans and intentions. The appearance of bootstrapping occurs when an intentional action is rational relative to the plan-constrained perspective internal to the agent’s framework of deliberation but irrational relative to the non-plan-constrained perspective external to the agent’s framework of deliberation.

Within the plan-constrained bounded practical reasoning of a planning agent, plans are formed and executed in various ways. Often an agent has a non-deliberative


intention. A non-deliberative intention is an intention that an agent has at t1 to A at t2 but that was formed at t0 and retained from t0 to t1 without reconsideration.\textsuperscript{171} In the absence of a problematic situation, the agent has no occasion to reconsider his settled course of action as embodied in his non-deliberative intention, and at the appropriate time simply executes the plan as it exists. “On the other hand, there are cases in which the agent deliberately weighs the pros and cons in favor and against several conflicting options that present themselves for choice.”\textsuperscript{172} These deliberations are prompted when an agent encounters a problematic situation, a situation in which one discovers that his or her beliefs at the time he or she formed the intention were mistaken to such an extent that they can no longer justify one’s intentions, that the circumstances surrounding the opportunity to execute one’s plans are significantly different from the circumstances one expected when forming the intention, or one’s values have changed between the time the intention was formed and the time an opportunity arises for its execution such that one no longer regards as justified the end toward which the intention is directed.\textsuperscript{173} These cases often require deliberative decision and the formation of a deliberative intention. A deliberative intention is an intention the agent has at t1 to A at t2 that is the result of present deliberation, \textit{i.e.}, it is an intention formed at t1.\textsuperscript{174} However, most deliberative decisions occur within the framework of non-deliberative plan execution, and so most examples of intellectual

\textsuperscript{171} Bratman, \textit{Intentions, Plans and Practical Reason}, 56.

\textsuperscript{172} Bratman, “Taking Plans Seriously,” 278.

\textsuperscript{173} Bratman, \textit{Intentions, Plans and Practical Reason}, 67.

\textsuperscript{174} Bratman, \textit{Intentions, Plans and Practical Reason}, 56.
activity are ‘hybrids.’ In the hybrid cases, the framework of an agent’s intentions and plans provides a background for the weighing of pros and cons in such a way that only some options are seen as relevant and acceptable.

In addition to intentions and plans, human beings also utilize policies in order to compensate for their bounded rationality. Personal policies are neither habits nor rules of thumb, but general intentions applicable to recurring circumstances in one’s life. Like intentions and plans, general policies help extend the influence of deliberation beyond the present, facilitate coordination, but also are useful as defeasible guides to life because “it may sometimes be easier to appreciate expectable consequences (both good and bad) of a general ways of acting in recurrent circumstances than to appreciate the expectable consequences of a single act.”

Some intentions are based on these sorts of general policies. These intentions are called policy-based intentions. Policy-based intentions are intentions formed at t1 to A at t2 on the basis of a general policy to A in circumstances C and a belief at t1 that circumstances C will obtain at t2. Policy-based intentions are not policies, but rather more specific intentions based on policies, and are, properly speaking, neither non-deliberative nor deliberative intentions. They are not non-deliberative because it is not a case of retaining a prior intention, but rather involves the specification of a policy by giving it a more specific content. It is not deliberative because it does not involve a full-blown weighing of pros and cons, but merely a specification of a prior policy.

In bounded practical reasoning involving policies, the requirement of coherence extends to policies as well. An agent must maintain coherence between a plan and a policy, and coherence within policies. The idea is that one’s policies can raise problems analogous to high-level future-directed intentions that need to be resolved in order to maintain coherence. “Normally, the need for coherence between policies and plans results in a need to settle on a specification of a general policy. I ensure such coherence by adding to my plans a policy-based intention to carry out my policy on an appropriate occasion.”

The framework-reasons view of the planning theory of human agency allows via policies for certain commitments to non-specified future action to play a role as inputs into practical reasoning. In ways analogous to future-directed intentions, personal policies allow an agent to coordinate, plan, and act in ways that are consistent with his or her overall commitments. Being able to regulate one’s behavior in this way is necessary in order to cope with human nature’s limitations. However, future-directed intentions, plans, and personal policies are not sufficient to regulate our behavior in the appropriate ways. Bratman recognizes that often our beliefs can conflict with our commitments. In such cases, we need to make choices.

Bratman wants to break from what he calls the standard model of decision theory that requires that all decisions, and that the operation of the will, be related to the choice of present options. As he says,

On the…planning theory, the ‘will’ is expressed also in the choices of (and intentions concerning) future

options…. On the model suggested by the planning theory, I can make a present choice in favor of a valued sequence of future actions or a valued policy to act in certain ways on certain occasions. Such a choice issues in an intention concerning future conduct, and this intention is, normally, causally responsible for specific, future actions.\footnote{179}

Bratman’s account of choice suggests that in addition to being planning agents, we are also agents who must view ourselves as temporally extended agents. Not only do we have futures, but our plans have futures as well. It is important to living out our commitments that our future selves at plan’s end will not regret what our present selves have done or failed to do at plan’s beginning.\footnote{180}

Bratman takes it as a fact of human experience that we see ourselves as temporally extended agents. How are we to understand the possibility that we are temporally extended agents? Bratman offers a strategy. He begins by appeal to the neo-Lockean attempt to argue that personal identity over time consists in “psychological strands of psychological connectedness” and continuity extending both into the past but also into the future.\footnote{181} He then assumes that the presence of these connections and continuities is the result of the presence and operation of relevant states and attitudes that play the appropriate roles in the agent’s psychology during his practical reasoning and action. These states and attitudes must support the


temporal organization of the agent. Intentions, plans, and policies play a large role in this temporal organization. However, Bratman does not think that these psychological states are sufficient. What other attitudes and states are required? In order to answer this question we must look at an additional concern of Bratman’s.

Bratman’s introduction of the concept of temporally extended agency raises the problem of understanding the relationship between planning agency and deeply responsible agency, or what he sometimes calls agential authority. The problem of agential authority is to describe how it is possible that human actions are not simply determined by one’s desires or pro-attitudes, but rather that agents determine, govern and direct their actions.

What we want to know is whether there is a kind of psychological functioning, a kind we can characterize without presupposing the very idea of agent determination of action, such that agent determination consists in such functioning…. S is the full-blown agent of X iff X is the issue of psychological functioning of type T…. The problem of agential authority is the problem of specifying T or, alternatively, arguing that no such account is available.\(^\text{182}\)

Bratman provides an account of deeply responsible agency that is based on the human capacity to engage in higher-order critical reflection. In this sense, he is following Harry Frankfurt. Harry Frankfurt has attempted to develop an account of responsible agency with an appeal to the concept of identification.\(^\text{183}\) The central idea of


Identification is that in order for an action to be regarded as fully one’s own, upon reflection and endorsement, one must identify with one’s own motivation. Identification is defined in terms of satisfaction, and Frankfurt says satisfaction entails “an absence of restlessness or resistance”\(^{184}\) with respect to one’s desires. However, Bratman argues that this view requires an agent to make a decision with respect to his or her desires.\(^{185}\) One must still decide whether to change one’s desires or to leave good enough alone. What is needed, according to Bratman, is an account of identification not merely in terms of Frankfurt’s notion of satisfaction, but in terms of a decision with which one is satisfied.

Unlike Frankfurt, Bratman will emphasize planning agency and argue that deeply responsible agency through identification is in part dependent on planning agency. Bratman appeals to Scanlon in an effort to understand what kind of decision is involved in identification. Scanlon has argued that sometimes in practical reasoning we make decisions concerning what to count as reasons for action.\(^{186}\) Bratman suggests that regarding a desire as external (as an outlaw) is to decide not to consider a particular desire as a reason for action (or as regarding its satisfaction as an


\(^{185}\) See Bratman, “Identification, Decision, and Treating as a Reason,” in *Faces of Intention: Selected Essays on Intention and Agency* (New York: Cambridge UP, 1999), 194-195. Notice that Bratman is insisting that identification involve satisfaction with one’s desires. Bratman never appears committed to a Humean theory of evaluation which this view of identification presupposes. Nonetheless, his arguments (almost) always presuppose it. I reject these Humean presuppositions.

\(^{186}\) See T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard UP, 1998). See also Rachel Cohon, “Internalism about Reasons for Action,” Pacific Philosophical Quarterly 74 (1993): 265-288; and Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard UP, 1990). I should point out that Scanlon’s basic claim about deciding to count something as a reason for action does not require deciding to count a desire as a reason for action. As I stated above, Bratman is presenting a desire-satisfaction account of foundational reasons for action. While his initial argument in *Intentions, Plans and Practical Reason* is agnostic about the normativity of desires (see 52-53), his arguments continually assume the normativity of desires. As I will argue in Chapter 4, a moral psychology of virtue cannot endorse a desire-satisfaction account of foundational reasons for action.
end to which one’s actions should be directed). Similarly, identification with a desire is a decision to regard that desire as reason-giving and to regard its satisfaction as an end toward which one’s actions are directed (given the coherence, consistency and stability of one’s overall intentional structure of future-directed intentions, plans and policies). “One treats one’s desires as reason giving when one treats it as setting an end that can to some extent justify means and/or preliminary steps.”\textsuperscript{187} However, identification involves more than this kind of decision. It also involves, as Frankfurt has suggested, satisfaction with the decision made. To be satisfied with a decision made is at least not to have “reached and retained a conflicting decision, intention, or policy concerning the treatment of one’s desire as reason-giving.”\textsuperscript{188}

I should pause to point out that I do not think that the moral psychology of character which I will be developing is committed to Bratman’s Humean theory of evaluation and practical reason. In fact, it is not clear that Bratman is committed to it. In Intentions, Plans and Practical Reason, Bratman twice expresses reservations about a Humean theory. For example in Chapter 2 he says, “one might argue that only desires that would remain after full rational reflection on available information provide [considerations that are directly relevant to the rationality of (an agent’s) actions.] Or one might urge, in an Aristotelian spirit, that only facts about what things are part of or a means to an objectively good human life can provide such direct considerations. These challenges are deep and important. But I shall put them

\textsuperscript{187} Bratman, “Identification, Decision, and Treating as a Reason,” 198.

\textsuperscript{188} Bratman, “Identification, Decision, and Treating as a Reason,” 201.
Later, in Chapter 4, Bratman says, “I am going to assume that the end relative to which we are to make such consequentialist assessments [of agent rationality] is the agent’s long-term interest in getting what she wants. Here, however, the reader should recall my caveat, in Chapter 2, concerning the limitations of my discussion. I continue to put to one side possible criticisms of an agent’s basic desires. You could say, if you wanted, that judgments of agent rationality are relative to the agent’s long-term interests in satisfying her rational desires, or her long-term interest in leading an objectively good life, or something similar.”

Bratman believes that his theory of the temporally extended planning theory of human agency does not depend on any particular view of how we ought to assess agent rationality.

I will return in Chapter 4 to the issue of Bratman’s appeal to a desire-satisfaction account of foundational reasons for action. At this point I simply want to mention that one can accept Bratman’s descriptive account of temporally extended planning agency, include his account of identification resulting from a decision, without accepting his Humean assumptions.

Let us now return to the question of what psychological states, in addition to intentions, plans, and policies, are required to support the temporal organization of a planning agent. Drawing on Frankfurt’s appeal to satisfaction and his own account of identification, Bratman suggests that self-governing policies may play the

---


191 A self-governing policy is a higher-order policy concerning the development, support and implementation of lower-order policies. A self-governing policy is, as the name implies, a policy directed toward the development of policies with which to regulate one’s conduct. See “Reflection, Planning, and Temporally Extended Agency,” 47 – 48. See also “Valuing and the Will,”
appropriate role in creating Lockean connections and continuity, provided one is satisfied with one’s self-governing policies. “For one to be satisfied with one’s self-governing policy, is …for that policy not to be challenged by one’s other self-governing policies” where policy P is challenged by policy P* when “P* is in conflict with P and, as a result, the presence of P* tends to undermine the role of P in supporting coordinating, Lockean ties.”\textsuperscript{192} Bratman’s suggestion is that an agent’s identification with a desire is determined by the endorsement or rejection by one’s self-governing policies that the satisfaction of this desire function as providing a justifying end in motivating efficacious practical reasoning. It is not just that the desire act as a reason for action, though it might, but also that it shape and direct further practical reasoning.

Self-governing policies, and so human planfulness, support temporally extended agency by organizing and unifying practical reasoning over time. Bratman argues that

Such self-governing policies have it as part of their organizing role to support a temporally extended pattern of functioning of relevant desires in ways that help constitute and support Lockean ties characteristic of our temporally extended agency. And that is why they have authority to speak for the agent.\textsuperscript{193}

\textit{Philosophical Perspectives 14: Action and Freedom} (2000): 249-265 where Bratman describes a self-governing policy as “higher-order policies that concern not only the motivation role of [an agent’s] desires, but also its treatment of its desires as reason-providing in motivationally effective deliberation. These are policies governing what weight, if any, to give to desired ends in motivationally effective deliberation” (258).


\textsuperscript{193} Bratman, “Reflection, Planning, and Temporally Extended Agency,” 50.
Bratman describes self-governing policies as playing a role similar to Charles Taylor’s discussion of efforts to articulate what is of “decisive importance.” Self-governing policies allow an agent to take a stand with respect to his array of motivations. Therefore, the explanatory power of reflective endorsement and strong reflectiveness arises from the role that future-directed intentions, and so policies, play in human agency.

All hierarchical theories must also deal with the problem of subjective authority. The problem is how are we to relate our having a pro-attitude toward X and our regarding X as having “normative authority in one’s motivationally effective deliberation.” In his attempt to solve the problem of subjective normative authority, Bratman asks what content is required in order for the intentions, plans, and policies and self-governing policies to provide the necessary Lockean continuities and connections to sustain personal identity. In answering this question, Bratman fully endorses an appeal to hierarchical structures, but a structure developed on the ideas of Korsgaard rather than Frankfurt:

At this point we draw on a more Korsgaardian idea: we appeal to higher-order attitudes concerning what desired ends to treat as justifying considerations in motivationally effective deliberation. Now, the hierarchical story I have been developing appeals to higher-order intentions, plans, and policies. So the idea would be to give these higher order attitudes content that concerns what desired ends to treat as justifying reasons. We try to solve the problem of subjective normative authority, and to say what constitutes the agent’s deliberation, by appeal to higher-order

---


intentions, plans, and policies about which desired ends to treat as reasons in motivationally effective deliberation.\textsuperscript{196}

Self-governing policies about which ends to treat as reasons for X provide the necessary connection between one’s mere pro-attitude toward X and one regarding X as having normative authority in his or her motivationally effective deliberation. In order for this account to work, Bratman must provide an explanation of how agents develop self-governing policies regarding which ends to treat as reasons. Bratman’s solution is to argue that human agency also requires the concept of valuing, that valuing involves more than merely wanting, and that valuing something cannot be reduced to \textit{merely} judging that it is good.\textsuperscript{197} To value something is to regard it as of decisive importance and the place it at the base of one’s framework of bounded practical reason.\textsuperscript{198}

With this description of the temporally extended planning theory of human agency in place, we are now in a position to articulate more specifically how this theory can help provide for the functional roles of the virtues in such a way that the psychological conditions of possibility for being virtuous can be satisfied. Before turning to that exercise, I want to present as a conclusion a table representing

\textsuperscript{196} Bratman, “Two Problems about Human Agency,” 321.

\textsuperscript{197} See Bratman, “Valuing and the Will,” especially 260.

\textsuperscript{198} Bratman is attempting to avoid an account of valuing that reduces valuing to either a desire or a judgment. Valuing is, strictly speaking, neither a judgment that something is good, nor a desire. Instead it is the result of the combination of a hierarchical structure of desires (a la Frankfurt) and a planning structure of intentions, plans and policies (a la Bratman). Valuing is a result of the “capacity to arrive at policies that express [an agent’s] commitment to being motivated by a desire by way of its treatment of that desire as providing, in deliberation, a justifying end for action” (Bratman, “Valuing and the Will,” 258) The policies at issue here are self-governing policies. The role of policies in this account is that the operation of a policy can shape what an agent desires and what she values.
Bratman’s Gricean creature construction\textsuperscript{199} examining a number of different models of action. Each model is adequate for one creature, but then it must be supplemented and extended in various ways as the creatures come to resemble human beings. During the construction, we can identify a number of core features of human agency. An adequate model of human agency must account for these core features. The result of this construction is the identification of a number of different conceptions of valuing. In addition, it allows us to see how complex the temporally extended planning theory of human agency is. It should be noted that the descriptions of the creatures do not follow exactly the discussion I have present above. However, many of the elements of the temporally extended planning theory of human agency, as I have presented it, can be seen below. Of particular interest for my purposes are the ways in which Creatures 4 – 8 differ from 1 – 3.

TABLE 3.1 BRATMAN’S GRICEAN CREATURE CONSTRUCTION

<table>
<thead>
<tr>
<th>CREATURE</th>
<th>ATTITUDES</th>
<th>CAPACITIES</th>
<th>VALUING</th>
<th>ACTION</th>
<th>PROBLEM</th>
<th>SOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creature 1</td>
<td>Beliefs</td>
<td>To desire and have beliefs</td>
<td>None</td>
<td>Moved by its strongest desire at the moment</td>
<td>Lacks reflection</td>
<td>Consider desires in light of beliefs</td>
</tr>
<tr>
<td></td>
<td>Desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creature 2</td>
<td>Beliefs</td>
<td>Consider desires in light of beliefs about experience, consequences, etc.</td>
<td>None</td>
<td>Moved by its strongest considered desire at the moment</td>
<td>Conflicting considered desires</td>
<td>Deliberation takes desire to be providing a justifying end</td>
</tr>
<tr>
<td></td>
<td>Desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Considered desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREATURE</td>
<td>ATTITUDES</td>
<td>CAPACITIES</td>
<td>VALUING</td>
<td>ACTION</td>
<td>PROBLEM</td>
<td>SOLUTION</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Creature 3</td>
<td>Beliefs</td>
<td>Normative deliberation in which a considered desire is thought to provide a justifying end</td>
<td>Values what it, after consideration, desires</td>
<td>Moved by what its takes to be its most justified considered desire at the moment</td>
<td>Rationally limited agents lack intra- and inter-personal coordination over time</td>
<td>Future-directed intentions, plans, policies concerning its actions; and acceptance in a context</td>
</tr>
<tr>
<td></td>
<td>Desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Considered desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Considered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creature 4</td>
<td>Beliefs</td>
<td>Settle on partial plans, policies, and accepts propositions when practical pressures require it</td>
<td>Same as above</td>
<td>Moved by what it takes to be its most justified desire given its intentions, plans, and policies concerning action</td>
<td>Temptation via changing desires and temporal discounting</td>
<td>Recognition of self as temporally extended agent and plan stability via the no-regret principle</td>
</tr>
<tr>
<td>200</td>
<td>Desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Considered desires</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intentions, plans, and policies concerning action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acceptance in a context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

200 Creature 4 is a planning agent.
## TABLE 3.1 BRATMAN’S GRICEAN CREATURE CONSTRUCTION

<table>
<thead>
<tr>
<th>CREATURE</th>
<th>ATTITUDES</th>
<th>CAPACITIES</th>
<th>VALUING</th>
<th>ACTION</th>
<th>PROBLEM</th>
<th>SOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creature 5</td>
<td>Beliefs Desires Considered desires Intentions, plans, and policies concerning action Acceptance in a context</td>
<td>Ability to anticipate future regret</td>
<td>Same as above</td>
<td>Moved by what it takes to be its most justified desire given its intentions, plans, policies and future self</td>
<td>Cross-temporal incoherence in which an agent’s intentions, plans, policies generate conflict over time</td>
<td>Reflective endorsement and strong reflectiveness</td>
</tr>
<tr>
<td>Creature 6</td>
<td>Belief Desires Considered desires Intentions, plans, and policies Acceptance in a context Higher-order desires</td>
<td>Capacity to arrive at higher-order desires (i.e.,) a will</td>
<td>Higher-order desires with which one identifies help shape what, after consideration, is desired</td>
<td>Moved by what it reflectively endorses as its most justified desire given its intentions, plans, policies and future self</td>
<td>Agential authority(^{201})</td>
<td>Higher-order self-governing policies concerning its desires support Lockean connections and continuities</td>
</tr>
</tbody>
</table>

\(^{201}\) Recall that the problem is to describe how it is possible that human actions are not simply determined by one’s desires, even one’s reflectively endorsed desires, but rather that agents determine, govern and direct their actions.
## TABLE 3.1 BRATMAN’S GRICEAN CREATURE CONSTRUCTION

<table>
<thead>
<tr>
<th>CREATURE</th>
<th>ATTITUDES</th>
<th>CAPACITIES</th>
<th>VALUING</th>
<th>ACTION</th>
<th>PROBLEM</th>
<th>SOLUTION</th>
</tr>
</thead>
</table>
| Creature 7 | Desires  
Beliefs  
Considered desires  
Intentions, plans, and policies concerning action  
Acceptance in a context  
Higher-order desires  
Higher-order policies concerning desires | Ability to take a stand with respect to the functional (motivational) roles of its desires | Agent values those higher-order desires which are in accord with its higher-order (desire-governing) policies | Moved by what its higher-order desires which accord with its higher-order (desire-governing) policies | Subjective normative authority | Self-governing policies provide a justifying end. |

---

202 Creature 7 is a temporally extended planning agent.

203 Recall that the problem is how are we to relate our having a pro-attitude toward our higher-order desires and our regarding our higher-order desires as having normative authority in one’s motivationally effective deliberation.
<table>
<thead>
<tr>
<th>CREATURE</th>
<th>ATTITUDES</th>
<th>CAPACITIES</th>
<th>VALUING</th>
<th>ACTION</th>
<th>PROBLEM</th>
<th>SOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creature 8</td>
<td>Desires Beliefs Considered desires Intentions, plans, and policies Acceptance in a context Higher-order desires Higher-order policies Self-governing policies</td>
<td>Capacity to arrive at self-governing policies and arrive at first-order desires reflecting the commitments incorporated in the self-governing policies</td>
<td>An agent values X when he or she has a desire for X and a self-governing policy in favor of treating that desire as providing an end that is justifying in motivationally effective deliberation</td>
<td>Moved by its higher-order desires which its self-governing (desire-governing) policies determine to provide a justifying end</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5 Conclusion

Creatures 1 - 3 are versions of the simplistic view of human agency that Bratman attributes to many philosophers of action, particularly those who advocate the desire-belief model of human agency. Bratman repeatedly stresses that while such an account may be a useful account of purposeful action, it is far too simplistic to account for human agency. Bratman’s *Intention, Plans and Practical Reasons* was an attempt to construct creature 4 (and 5) while *Faces of Intention* contains essays that attempt to construct creatures 5 and 6 (and 7). His post *Faces of Intention* work, which has focused on hierarchical theories, has attempted to construct creatures 7 and 8. Creature 8 represents Bratman’s latest effort to articulate the nature of human agency as temporally extended planning agency. This account is, I think, a fruitful place to look for resources to meet Anscombe’s Demand.
CHAPTER 4

THE PSYCHOLOGICAL CONDITIONS OF POSSIBILITY FOR BEING VIRTUOUS

4.1 Introduction

Let’s take stock. I began with the claim that being virtuous involves the possession of character traits that provide an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability to act reliably for, or because of, those reasons. I also suggested that there are two components to a full understanding of the possibility of being virtuous, a normative component and a psychological component. We are justified in insisting on the existence of these two components because it is possible that while objectively appropriate reasons for action do exist, human beings are psychologically incapable of recognizing and acting for those reasons. This psychological incapacity might even hold under conditions in which a human being has full knowledge of these reasons. On the other hand, it is also possible that while human beings are psychologically capable of consistently recognizing and reliably acting for objectively appropriate reasons for action were they to exist, it is still unfortunately the case that objectively appropriate reasons for action in fact do not exist. Given
these possibilities, I insisted for the sake of argument on dealing separately with the normative and the psychological conditions of possibility for being virtuous.

The argument of Chapter 2 was a rejection of a particular understanding of the normative condition. According to the view I rejected, agent-basing, the evaluation of action is entirely dependent on the motivations and inner life of the virtuous agent. That is to say, when one evaluates an action as right, one is depending on an evaluation of the agent who performs the action. I argued against this view, concluding that we have good reason to believe that we must have a good-based account of right action. I left for another time the task of articulating a substantive positive account of the normative conditions of possibility for being virtuous. Articulating a defense of the objectivity and appropriateness of certain potential reasons for action is the work of an ethical theory, any defense of which I have avoided. I recognize that only certain ethical theories are consistent with the positions defended in this dissertation. I also recognize that some rational defense of one of these theories is required to complete a defense of virtue ethics. Nonetheless, it seems clear to me, as it did to Anscombe, that a defensible moral psychology of character is necessary for, if not a prerequisite to, an adequate defense of virtue ethics.

It does not follow from a good-based account of right action that the source of an agent having objectively appropriate reasons for action is not his or her traits of character. I want to argue that it is an agent’s character that allows him or her to correctly respond to reasons for actions which are objectively appropriate. So while there must exist objectively appropriate good-based reasons for action in order for an
account of virtue ethics to be vindicated, there must also exist a psychological apparatus that allows human beings to react correctly to those reasons. So, how is it possible for human agent’s to recognize and act based on these reasons?

According to the account of being virtuous developed here, being virtuous involves the possession of character, which provides an agent with the capacity consistently to recognize objectively appropriate reasons for action as well as the ability to act reliably for, or because of, those reasons. This understanding of being virtuous presupposes that despite all the limitations with which human beings are saddled, it is nonetheless possible for human beings to perform these tasks. I have been calling the truth of these presuppositions the psychological conditions of possibility for being virtuous. My primary task in this Chapter is to attempt to make good the claim that the temporally extended planning theory of human agency provides the resources for an account of the psychological conditions of possibility for being virtuous. If I can provide an account of how it is possible for human beings to have these capacities and abilities, then I will have provided a potential aid to the virtue theorist who wants to defend virtue ethics as a coherent moral theory.

Before moving on to this important task, I want to address a potential objection. The objection has to do with Bratman’s own account of practical rationality and the fact that his views on this issue are inconsistent with a good-based account of objectively appropriate reasons for action.

4.2 A Good-based Foundational Account of Reasons for Action

It might be supposed that since the temporally extended planning theory of human agency as presented by Michael Bratman has a normative account of practical
rationality which does not depend on a good-based account of reasons for action, it follows that the planning theory is inconsistent with my account in Chapter 2 of the normative conditions of possibility for being virtuous in that on Bratman’s account objectively appropriate reasons for action are not good-based. In the last chapter I avoided a discussion of Bratman’s normative account of practical rationality in order to distill what I think is a valuable descriptive account of bounded human rationality. Unfortunately, as a result of concentrating on this descriptive account, I have left room for a possible misunderstanding of the position being outlined here. Therefore, it is appropriate to address Bratman’s normative account of practical rationality.

Bratman’s normative account of practical reason is grounded in the idea that rational action satisfies desires. Bratman suggests and weakly endorses standards for the assessment of agent rationality derived from the pragmatic concern of agents’ long-term interests in getting what they want. The type of assessment of agent rationality that Bratman finds most plausible is consequentialist in nature. “We require (or so, anyway, I am here assuming) that the agent’s relevant habits, dispositions, and ways of arriving at decisions come up to a certain level of effectiveness in their expected impact on the agent’s long-term interests in getting what she wants.” Bratman is very careful to explain that while he thinks that this approach to the assessment of agent rationality is the most plausible, it is not the only option. Furthermore, he does think that his temporally extended planning theory of human agency is compatible with options such as assessments in terms of an agent’s

long-term interest in satisfying her *rational* desires or assessments in terms of an agent’s long-term interest in being an objectively good person. However, he does not offer any argument for his pragmatic consequentialist approach over these other options. Nor does he offer any argument for his restriction of possible options to methods of assessment which make reference to the agent’s interests. Standards of assessment could make reference to an agent satisfying his or her rational desires or to an agent being an objectively good person.

On Bratman’s view, then, the ultimate rationale for having intentions in the first place is to facilitate human beings of bounded rationality in the satisfaction of (rational) desire. Therefore, although intentions, plans and policies provide reasons for action in a secondary sense, the primary reasons for action are preferences and desires. “The expected satisfaction of (rational) desire provides reasons for action that are basic in a way in which the framework reasons provided by intentions are not.” The most basic reasons for action are preferences. Rationality is assessed, in part, in terms of an agent’s long-term interests in getting what he or she wants.

---


206 Notice that this claim makes no explicit reference to the interests of the agent in satisfying his or her desires. Recall that Bratman’s discussion of the human need for future-directed intentions and plans relied on an appeal to human limitations and human beings’ ability to compensate through the formation of intentions and plans. Therefore, the rationale of which he speaks in this context is really an explanation of why human beings have the psychological capacities they have. It is unclear how Bratman moves from this explanatory claim about the development of the human capacity to plan to a rationale or justification for the particular intentions or plans a particular agent has at a particular time. While the development of the capacity for adult humans to digest lactose provided a selective advantage in those populations which domesticated milk producing animals, an appeal to selective advantage does not necessarily provide a rationale or justification for the drinking of milk by a particular individual at a particular time.

However, Bratman also rejects the idea that practical rationality is solely a matter of rational calculation in terms of desire or preference satisfaction. Instead, he wants to supplement a desire or preference satisfaction account with an account of the norms associated with bounded practical reasoning, norms such as internal consistency, cognitive consistency, and positive means-end coherence. An assessment of agent rationality in terms of mere desire satisfaction is always an assessment made from a perspective external to the agent’s deliberation. That is, such assessment is non-plan-constrained. Since bounded human rationality is dependent on the formation and utilization of future-directed intentions, plans and policies, a complete account of agent rationality must incorporate an assessment of the coherence and consistency of an agent’s deliberation relative to the deliberative framework within which he or she is operating. To achieve this more complete assessment of agent rationality requires determining whether or not the deliberation, reconsideration, and retention of an agent’s intentions, plans, and policies, as well as the underlying habits and dispositions governing an agent’s reconsiderations meet an appropriate standard of rationality. Despite this apparent complication to the account, it is still the case that in making a complete assessment of agent rationality, the reconsideration of intentions, plans and policies, as well as the habits and dispositions governing an agent’s reconsiderations are assessed in terms of how well they facilitate desire-satisfaction.

Given this framework of assessment, Bratman proposes a two-stage process of assessing agent rationality. First, intentional action is assessed in terms of present-directed intentions. That is, one assesses whether or not an action is the sort of action
that will fulfill or satisfy the desire which gave rise to a present-directed intention. If an intentional action is rational for an agent in this sense, then one must ask whether or not it is rational of an agent to have the present-directed intentions that he or she has. According to Bratman, the first stage of assessment is made according to the intention-action principle which states that “if it is rational of S to have a present-directed intention to A, and S successfully executes this intention and thereby intentionally As, then it is rational for S to A.”

The application of the intention-action principle requires that its antecedent be fulfilled. We need to know when it is rational of S to have a present-directed intention to A. Or, more generally, under what conditions is it rational of S to intend at t1 to A at t1?

To answer this question, Bratman first distinguishes three different kinds of intentions. Deliberative intentions are intentions the agent has at t1 to A at t2 that are the result of present deliberation, i.e., intentions formed at t1. Non-deliberative intentions are intentions that an agent has at t1 to A at t2 but that were formed at t0 and that have been retained from t0 to t1 without reconsideration. Policy-based intentions are intentions formed at t1 to A at t2 on the basis of a general policy to A in circumstances C and a belief at t1 that circumstances C will obtain at t2.

In answering the question of when it is rational of S to have a present-directed intention, each of these kinds of intentions must be addressed.

---


209 Bratman offers an important complication to his idea of policies and policy-based intentions which I have omitted here. See his “Intentions and Personal Policies,” *Philosophical Perspectives 3: Philosophy of Mind and Action Theory* (1989): 441-469.
Non-deliberative intentions are subject to various forms of reconsideration. Nonreflective (non)reconsideration occurs when an agent just (non)reconsiders without engaging in any deliberation concerning whether or not to (non)reconsider. Nonreflective (non)reconsideration is the result of certain underlying habits, skills, and dispositions had by the agent. Deliberative (non)reconsideration is the result of present deliberation in which an agent weighs desire-belief reasons for and against reconsidering a prior intention. Policy-based (non)reconsideration, like deliberative and unlike nonreflective (non)reconsideration, involves the agent in explicitly addressing the question of whether to reconsider a prior intention, but unlike deliberative (non)reconsideration, in answering the question the agent appeals to a general policy regarding reconsideration in particular circumstances rather than to desire-belief reasons.

The nonreflective (non)reconsideration case is the case central to Bratman’s account of when it is rational of an agent to have a particular present-directed intention. His accounts of the other two cases will depend in large part on an understanding of this case. Bratman’s account of the assessment of nonreflective (non)reconsideration is a two-tier system. First, nonreflective (non)reconsideration of a prior intention is rational of S if it is the manifestation of general habits of (non)reconsideration that are reasonable of S to have. Second, habits of reconsideration are reasonable of S to have if they have a sufficiently positive expected impact on the long-term prospects of (rational) desire satisfaction for normal agents of our kind.
Reasonable habits of (non)reconsideration promote an appropriate level of stability in an agent’s intention as required for inter- and intra-personal coordination in rationally bounded agents like ourselves while also recognizing that (rational) desire satisfaction is sometimes promoted by de-stabilizing a prior intention in order to reflectively examine one’s life. Given our nature as temporally extended agents and given the nature of temporally extended deliberation, Bratman proposes that non-deliberative rationality be assessed according to the historical principle of non-deliberative rationality:

\[
\text{It is rational of S at } t_1 \text{ to intend to } A \text{ at } t_2 \text{ just in case } \\
1) \text{ it was rational of S at } t_0 \text{ to form this intention, and } 2) \text{ it was rational of S from } t_0 \text{ to } t_1 \text{ not to reconsider this intention.}
\]

On the historical principle of non-deliberative rationality, present rational non-reconsideration does not guarantee the present rationality of S in intending. Rather, rational non-reconsideration acts as a kind of rational link, transmitting earlier rationality to a later time. It is true that it may be rational of an agent non-deliberatively to intend to A even if an evaluation from an external perspective might determine that such an intention is not recommended in light of the agent’s long-term interest in satisfying his or her desires. For this case to hold, it must have been rational of the agent not only to not reconsider the intention but also rational to have formed it in the first place. So, on the historical approach the rationality of an agent in the present depends on past reasons, past deliberation, and past and present (non)reconsideration.

\footnote{Bratman, \textit{Intentions, Plans and Practical Reason}, 80}
Bratman’s historical principle of non-deliberative rationality requires an account of deliberative rationality in intending due to condition (1). Bratman suggests the historical principle of deliberative rationality:

If S at t1 forms the intention to A at t2 on the basis of deliberation at t1, then it is rational of S at t1 to intend to A at t2 iff: 1) for those intentions of S’s that play a direct role as a background of S’s deliberation, it is rational of S at t1 so to intend; and 2) S reasonably supposes that A is at least as well supported by his [framework] reasons for action as its relevant, admissible alternatives. 211

Bratman concludes that his historical approach to agent rationality allows for the maintenance of a distinction between an internal and an external perspective and corresponding assessments of the rationality of an agent in having a particular intention. He stresses repeatedly that these assessments can differ in either direction. For example, it is not rational (as opposed to irrational) of S to intend to B in order to A when it is irrational of S to intend to A. The account can also claim that while it is not rational of S to intend to B in order to A, it may be rational of S not to reconsider his intention to B in order to A. Also, from an internal plan-constrained perspective, given S’s intention to B, the account can admit that it is rational of S to intend to B.

With regard to divergence of assessments of S’s actions, as opposed to assessments of S, notice that while it might not be rational of S to intend to B in order to A, B-ing might still be recommended from an external perspective. Also, it might be rational of S to intend to B in order to A while B-ing would not be recommended from an external perspective.

211 Bratman, Intentions, Plans and Practical Reason, 85.
In addition to a principle covering non-deliberative and deliberative intentions, Bratman offers the historical principle of policy-based rationality. Some intentions are based on general policies. Policy-based intentions are not policies, but rather more specific intentions based on policies, and are, properly speaking, neither non-deliberative nor deliberative intentions. They are not non-deliberative because it is not a case of retaining a prior intention, but rather involves the specification of a policy by giving it a more specific content in the form of an intention. It is not deliberative because it does not involve a full-blown weighing of pros and cons, but merely a specification of a prior policy. Since policy-based intentions differ from the other intentions examined thus far, they require their own treatment. Bratman wants a principle for determining when it is rational of S to have a policy-based intention at t1 to A at t2. A historical approach to this question requires a principle that includes the rationality of blocking or not blocking the application of the policy. This leads him to develop the historical principle of policy-based rationality:

\[
\text{It is rational of S at t1 to have the policy-based intention to A at t2 iff: a) it was rational of S at t0 to form the general intention to A when C; and b) it was rational of S at t0 to t1 not to reconsider this general intention; and c) it was rational of S not to block the application of his general intention to this particular case.}^{212}
\]

What is common to each of these historical principles of rationality is that the rationality of the formation of an intention is dependent on an agent’s long-term interests in getting what he or she wants. In addition, the rationality of retaining or

\[\text{212 Bratman, Intentions, Plans and Practical Reason, 91.}\]
blocking an intention is initially assessed according to norms of consistency and coherence, while the norms of consistency and coherence are justified by reference to human beings’ long-term interests in getting what they want.

I have provided an extensive description of Bratman’s normative account of agent rationality. I have done so for two reasons: First, to illustrate the pattern in his thinking with respect to the ultimate foundation of normative assessments of agent rationality in desire satisfaction; Second, to provide a context for an alternative possible grounding of agent rationality in terms of a good-based account of objectively appropriate reasons for action.

The pattern of justification in Bratman’s theory of agent rationality is a pattern of pushing back the appeal to desire satisfaction to some foundational stage, while assessing the more proximate stages of human deliberation with an appeal to standards of consistency and coherence appropriate to the deliberative framework within which the agent is operating. And while Bratman provides arguments for why, as rationally bounded creatures, human beings are dependent on standards of coherence and consistency, he never provides an argument for the role of desire satisfaction as foundational in assessments of agent rationality. It is therefore possible to appropriate both Bratman’s descriptive account of temporally extended human agency as described in Chapter 3 and his normative account of bounded rationality with respect to the dependence of human beings on standards of consistency and coherence as described in his historical principles of rationality,
while rejecting his foundational account of agent rationality.\footnote{213} With respect to the features of Bratman’s view with which I am interested, nothing of importance hinges on Bratman’s foundational account of agent rationality, and it can be replaced with a good-based account of reasons for action.

On a good-based account of reasons for action, a reason for action is a particular fact about the world, the situation, or the individuals involved, and not, as on Bratman’s foundational account, an interest in satisfying a desire or a preference. We need not follow Bratman in claiming that “the expected satisfaction of (rational) desire provides reasons for action that are basic in a way in which the framework reasons provided by intentions are not.”\footnote{214} Instead, it is consistent with the parts of Bratman’s view in which I am interested to argue that proper relationships between an action and the good provide reasons for action that are more basic in a way in which the framework reasons provided by intentions, plans and policies are not. Given what Bratman has said about the historical principles of rationality, any foundational account of reasons for action is consistent with an account of framework reasons for action.\footnote{215}

Bratman’s un-argued adoption of a desire-satisfaction foundational account of reasons for action is not limited to his explicit discussion of agent rationality. His discussion of the problems of agential authority and subjective authority also appeals to an undefended desire-satisfaction account. Recall that the problem of agential

\footnote{213} See Bratman, \textit{Intentions, Plans and Practical Reason}, 52. Bratman only weakly endorses his pragmatic consequentialist account of agent rationality.


\footnote{215} Bratman admits as much. See \textit{Intention, Plans, and Practical Reason}, 22 and 52.
authority is to describe how it is possible that human actions are not determined by an agent’s desires or pro-attitudes, but rather that an agent governs and directs her own actions. According to Bratman’s solution, an agent is said to direct her own action when her self-governing policies have as their content that the satisfaction of a particular desire function as a justifying end in motivationally efficacious deliberation. That is, self-governing policies allow an agent to take a stand with respect to his or her array of motivations, motivations which Bratman confines to desire satisfaction. The ability of an agent to take a stand with respect to his or her array of motivations plays a functional role in supporting Lockean ties characteristic of our temporally extended agency. By connecting our present motivations with decisions concerning our past motivations, human agents are capable of determining, governing, and directing present action on the basis of past deliberation.

However, the commitment characteristic of a self-governing policy to regard an end as a justifying end need not be a commitment to the satisfaction of one’s own desires. A self-governing policy is, as the name implies, a policy directed toward the development of policies with which to regulate one’s conduct. According to Bratman, self-governing policies are “higher-order policies that concern not only the motivation role of [an agent’s] desires, but also [an agent’s] treatment of [her] desires as reason-providing in motivationally effective deliberation. These are policies governing what weight, if any, to give to desired ends in motivationally effective deliberation.”

Given what Bratman says here, one might conclude that the object of an agent’s self-governing policies must be a desired end. In other words, a self-

governing policy says that one is to be motivated by a certain desire and regard the satisfaction of that desire as justifying.

However, anything which an agent takes to be of decisive importance can operate as an object of a self-governing policy. The solution to the problem of agential authority on a good-based account of reasons for action is not an appeal to desire satisfaction, but an appeal to commitment. An agent, who is committed to some object as being of genuine value, regards this object as a justifying end in motivationally efficacious practical deliberation. Recall our prior discussion of commitment in Chapter 3. There I emphasized four central characteristics of commitment: its temporal stability; its action-guiding force; its relation to self-understanding and identity; and its practical necessity. An agent’s commitment to something valued is captured in a self-governing policy which operates as a basic framework reason to which, in an ideal case, all other framework reasons making up the agent’s deliberative framework are consistently and coherently related. (If it is the case that the good is made up of constitutive parts and constitutive means, then each of these parts and means provides basic framework reasons.) Within the range of basic framework reasons resulting from the decision to regard certain valuable things as justifying ends, an agent makes decisions from within his or her deliberative framework concerning what to count as more specific reasons for action. It is possible that from within this deliberative framework an agent might be able to maintain consistency and coherence while also regarding the satisfaction of a desire as a reason for action. The satisfaction of desire is often good and perfectly
consistent with regarding something other than desire satisfaction as a basic justifying end.

As with the case of the historical principles of rationality, nothing of importance hinges on Bratman’s un-argued insistence on appealing to desire satisfaction as the object of an agent’s self-governing policies. All that is required to solve the problem of agential authority is an appeal to justifying ends as embodied in self-governing policies. In fact, Bratman’s solution to the problem of agential authority in no way depends on the content of self-governing policies. His solution to the problem is a functionalist solution – what is required to solve the problem is something to play the role of transferring the responsibility for having made decisions in the past to actions in the present. Self-governing policies do this, irrespective of their content.

The solution to the problem of subjective authority on a good-based account of reasons for action is similar to the solution to the problem of agential authority. Recall that the problem of subjective authority is how to relate our having a pro-attitude toward X and our regarding X as having normative authority in our motivationally efficacious deliberation. Bratman’s solution to this problem, unlike the solution to the problem of agential authority, does require an account of the content of an agent’s self-governing policies. While Bratman initially argues that the content of the self-governing policy must concern what desires to treat as justifying reasons, he later admits that the content of self-governing policies must concern what an agent values. To value something is to regard it as of decisive importance and to place it at the base on one’s framework of bounded rationality. Bratman asks, “Does
this mean that, strictly speaking, what one values (in this sense) is itself a desire – not honesty, say, but a desire for honesty? No it does not. What I value in the present case is honesty; but, on the theory, my valuing honesty in part consists in certain higher-order self-governing policies” concerning the roles of my desires in my deliberation.217 Since it is possible for an agent to value things other than the satisfaction of his or her own desires, there is no necessary connection between Bratman’s solution to the problem of subjective authority and his appeal to a foundational account of reasons for action in terms of desire satisfaction. All that is required for Bratman’s solution is the presence of the practical necessity characteristic of a genuine commitment to what one values. In light of one’s commitment to some valued end, one constructs self-governing policies to regulate the operation of one’s desires in one’s deliberation. By reference to the object of value, one is able to determine whether a framework reason or a first-order reason is a good reason for action. Without reference to honesty, an agent cannot determine with lower-order reasons for action to regard as good reasons for action. It is by appeal to the object of value that an agent can determine which of his or her lower order reasons for action are good reasons and which are bad. In a properly functioning deliberative structure, only good reasons for action have normative authority in our motivationally efficacious deliberation.

For the sake of illustration, consider neo-Aristotelianism, a view I characterized in Chapter 1 as holding that virtue is a constitutive means to and a constitutive part of the human good or telos, and that the evaluation of human action

is entirely derived from and dependent on what we have to say about the human good. Neo-Aristotelianism holds that an action’s positive relationship to the human telos is an objectively appropriate reason for action. This is a substantive normative claim which requires an ethical theory to substantiate it. However, if such an ethical theory could be provided, it would follow that the sort of reasons for action defended in the theory would be more basic than the framework reasons provided by intentions, plans and policies. However, given the facts of human bounded practical rationality, direct appeal to these reasons for action would be rare, except in cases requiring significant self-reflection due to the occurrence of (morally) problematic situations which threatened the stability of an agent’s deliberative framework. Instead, since a properly educated agent values the human telos (even if he or she could not provide a complete articulation of what the human telos is), her self-governing policies embody commitments to objects of genuine value which operate as justifying ends, are regarded as of decisive importance and are placed at the base of her framework of bounded practical reasoning.  

So, if 1) we assume, as I am, that the possibility of being virtuous requires the existence of objectively appropriate reasons for action, 2) it is true that these reasons for action must be good-based, and 3) it is the case that an agent’s desires, no matter their hierarchical order, are agent-based, it follows that objectively appropriate reasons for action cannot be confined to the prospect of desire satisfaction. It also  

---

218 One might object at this point that an appeal to the properly educated agent undoes all the work the historical principles of rationality do in supporting the idea of Lockean ties characteristic of our temporally extended agency. However, I think there is an argument to be made that Bratman’s historical principles of rationality can be extended to support a conception of inherited rationality. If one can then inherit rationally justified commitments and then rationally non-reconsider those commitments, it follows that a proper education can provide a foundation for the solutions to the problems of agential and subjective authority.
follows that an account of being virtuous cannot be built on Bratman’s foundational account of agent rationality. If we then also accept that there is no necessary connection between the temporally extended planning theory’s descriptive account and Bratman’s un-argued adoption of a desire satisfaction foundational account of agent rationality, it follows that it is coherent and consistent to jettison Bratman’s foundational account of agent rationality in favor of a good-based account of objectively appropriate reasons for action. What we retain from Bratman’s account, namely, the descriptive account and normative standards regulating an agent’s deliberative framework, is an account of a mental apparatus which is easily adapted to provide an account of character.

4.3 Toward a Moral Psychology of Character

Call the mental apparatus that fulfills the functional roles of the virtues ‘character.’ An individual who possesses the mental apparatus necessary for virtue is said to ‘have character,’ while those who lack this structure are said to ‘lack character.’ A trait of character is a specific description of an agent’s character or lack of character with respect of some specific end or good. Character, then, is not, properly speaking, a mental state; it is the result of stable and temporally extended relationships between a host of mental states, particularly relationships between desires, beliefs, acceptances, intentions, plans, policies and self-governing policies. The task of a moral psychology of character is to identify and describe these relationships.

By and large, the work of identifying and describing these relationships has already been done. Character is the result of a deliberative framework characteristic
of Creature 8 as described in Table 3.1 at the end of Chapter 3. Recall that Creature 8 is an evolution out of prior creatures. Therefore, Creature 8 has the following capacities: 1) to desire and to have beliefs, 2) to consider its desires in terms of past experience and expectations of future consequences, 3) to choose between various possible courses of action by weighing the possibilities in terms of their expected future consequences 4) to coordinate its actions over time via the formation of intentions, plans, and policies as well as acceptances, 5) to recognize itself as a temporally extended creature and anticipate future regret, 6) to arrive at higher order attitudes concerning the priority to be given to the range of plan-constrained lower order reasons for action, 7) to take a stand with respect to the decisive importance of particular reasons for action via the formation of self-governing policies, and finally, 8) to have first-order reasons for action reflecting the commitments incorporated in its self-governing policies.\(^{219}\)

An agent of character, then, is an agent who has commitments to ends or goods which are captured in his or her self-governing policies and provide for stable and temporally extended foundational framework reasons for action. The objects of these self-governing policies serve as foundational framework reasons for action. As a form of intention, self-governing policies involve volitional commitment, reasoning-centered commitment, while also being subject to internal consistency, cognitive consistency, and positive means-end coherence. What is unique about self-governing policies is the strength of the commitment involved. Intentions, plans and policies all resist reconsideration (due to the inertia characteristic of reasoning-

\(^{219}\) See Bratman, “Valuing and the Will,” 259-261.
centered commitment). Reconsideration occurs when an agent is faced with a problematic situation. In the case of intentions, a problematic situation can arise anytime it comes to an agent’s attention that his or her intentions lack consistency or coherence given his or her other intentions. However, in the context of a plan, the intentions of an agent are plan-constrained. Recall that plans pose problems for deliberation about further intentions and actions while simultaneously limiting the range of acceptable options. In cases where an intention and a plan conflict, the norms governing bounded rationality require that the conflict be resolved in favor of the plan, unless the plan itself is reconsidered.

Typically plans are reconsidered for reasons similar to the reasons for reconsidering intentions; namely, the plan is either inconsistent with other plans the agent has or the plan destabilizes the coherence of the agent’s deliberative framework. In either case, the resolution of the conflict is typically made in terms of some higher order attitude, such as a personal policy or a self-governing policy. Plans are typically directed toward something which an agent regards as good. However, merely regarding an object as good does not place any rational pressure on an agent to pursue that object in those cases where the object of the plan conflicts with the objects of an agent’s self-governing policies. The objects of an agent’s self-governing policies are the objects of the agent’s deepest commitments. It is these objects which the agent regards as of decisive importance and gives to them at least implicit deliberative priority.

In those cases where a plan comes into conflict with an agent’s self-governing policies, the norms governing bounded rationality require that the conflict be resolved
in favor of the self-governing policy, unless the self-governing policy is reconsidered. The reconsideration of a self-governing policy is an exercise in subjecting one’s basic and deepest commitments to a radical form of critique. The reconsideration of a self-governing policy is not the result of a problematic situation in the sense of an agent’s awareness of inconsistency or incoherence, though it might be sparked by such awareness. Rather the problematic situation arises out of a destabilization of the foundations of an agent’s deliberative framework. What the agent took to be of decisive importance, as a guiding and ordering principle of his or her life, is at risk of being rejected. Unfortunately, there is no higher order attitude to which one can appeal in order to resolve the problematic situation. It is perhaps better to call such problematic situations ‘moral crises.’

A moral crisis must be faced with reflective deliberation concerning the very foundations of one’s projects, practices, institutions, traditions, and significant relationships. As Helen Haste has argued, “In the resolution of this crisis the individual gains new moral (or political) perspectives, a new sense of personal agency, and a shift in perceived role and responsibilities.”\footnote{Helen Haste, “Moral Responsibility and Moral Commitment” in The Moral Domain: Essays in the Ongoing Discussion between Philosophy and the Social Sciences, ed. Thomas E. Wren (Cambridge, MA: MIT Press, 1990), 316.} The resolution of a moral crisis results in a moral conversion, the discovery of alternative foundational commitments which resolve the problems that precipitated the original crisis. By resolving a moral crisis in a manner which solves the problematic situation, stability is reintroduced into the agent’s deliberative framework.
I suggest using ‘character’ to capture the stability of the deliberative framework of Creature 8. Creature 8 possesses the psychological apparatus of an agent capable of having the kinds of commitments which can provide for the possibility that human psychology can fulfill the functional roles of the virtues. To say an agent has character has two related meanings. Often, ‘having character’ is a normative claim about the agent; it is a claim that the agent has a good character. We typically mean that an agent has deeply held commitments of an appropriate type, and that the agent has ‘integrity’ in the sense that his or her beliefs, intentions, emotions, and actions are consistent and coherent with respect to those commitments. However, it also makes sense to say that an agent has character where what we mean is that the agent has deeply held commitments of an inappropriate type, and that the agent has ‘integrity’ in the sense that his or her beliefs, emotions, and actions are consistent and coherent with respect to those commitments. So we might say that Hitler or Stalin were men of character, but only in this latter sense. Therefore, it appears that one sense of ‘having character’ refers to the consistency and coherence of an agent’s beliefs, desires, intentions and actions with respect to his or her commitments. A second sense of ‘having character’ refers to the nature of an agent’s commitments.221

221 One might object that this account of ‘having character’ provides too narrow a view of character. After all, it seems possible that an individual agent might refuse to consider consistency and coherence. Since human activity does not require consistency (or even coherence in some cases) it seems possible that an agent could act at any given time on the basis of his strongest desire without regard to the consistency or coherence of his actions over time. In this circumstance, it makes sense to talk about the agent having a character. In this case we might say that he has an intemperate character. However, this sense of ‘having character’ is different from the one I have been using. The sense of ‘having character’ as used in the objection implies that we have made an assessment of the sort of person the agent is. We have judged that he is intemperate. Therefore we can say he has a character, namely an intemperate character.

I have no objection to this use of ‘character.’ However, I think that when we say an intemperate agent has character, what we really mean is that he is lacking something which results in us describing him as intemperate. (See below my discussion of lacking character).
Alternatively, we often say that persons such as Hitler or Stalin lack character. However, ‘lacking character’ also has two senses. In the first sense, in the sense it is used with respect to Hitler and Stalin, what is meant is that the agent does not have a good character. That is, ‘lacking character’ in this sense refers to the nature of the agent’s commitments. In another sense, ‘lacking character’ means lacking consistency and coherence in one’s desires, beliefs, intentions, and actions with respect to one’s commitments (or as a result of one’s lack of commitment). In this sense of ‘lacking character’ what is meant is that an agent lacks the psychological apparatus characteristic of Creature 8. To say someone lacks character in this sense is not itself a normative assessment, but rather a psychological description. It is a description of a lacking. However, any number of missing aspects of the psychological apparatus may account for an agent’s lack of character.

There are, then, two sense of ‘having character’ and two senses of ‘lacking character.’ In each case one sense is primarily a psychological description while the other sense is primarily a normative assessment of the nature of the objects of an agent’s commitments. Often it is the case that when we use either of these terms, we mean it in both senses. So, perhaps the most common usage of ‘having character’ is both psychological and normative. However, it need not be, as the cases of Hitler and Stalin make clear. On the other hand, the most common uses of ‘lacking character’ appear to be psychological. The uses of ‘lacking character’ are often used to describe individuals who lack integrity, who lack the ability to stick with their intentions and plans, who succumb to temporary preference changes, temporal discounting, temptation, or social pressure. In other words, ‘lacking character’ describes the
psychological status of an individual who is unable to remain temporally consistent and coherent with respect to his intentions.

We often refer to the possession of various aspects of the psychological apparatus of character in different ways. These various ways of talking about the possession of or lack of character are typically conducted in terms of ‘traits of character.’ To discuss a trait of character is to specify character in some way, either with respect to a particular kind of ability or with respect to a certain kind of object. For example, we might talk about an agent’s traits of character with respect to his or her ability to resist temptation, or to risk his or her life for the sake of a foundational commitment. Alternatively, we also speak of an agent’s ability to act toward achieving particular goods for the sake of another, or to recognize the truth concerning objects of genuine value. In these ways we may speak of temperance, courage, justice and prudence. As is the case with discussions of character, ‘trait of character’ also appears to have a primarily normative sense and a primarily descriptive sense. Sometimes when we talk about traits of character we are talking about the nature of the particular objects of the commitments of an agent, and sometimes we are talking about the fact that the agent has or does not maintain consistency and coherence with respect to his commitments (or because he has no commitments at all). The latter use of traits of character suggests that there is also a positive and a negative descriptive sense of ‘trait of character.’ Sometimes ‘trait of character’ is a description of an ability on the part of the agent. Sometimes ‘trait of character’ is a description of an inability on the part of the agent. When talking about a person of character, we can talk about his or her abilities or inabilities to act
consistently and coherently in certain areas of his or her life and with respect to certain goods. So, for example, we might talk about a person’s traits of character with respect to his or her dealings with other persons. Some of these ascriptions of traits of character are primarily normative, such as ascribing to someone honesty or dishonesty. Other ascriptions of a trait of character are primarily descriptive, such as ascribing to someone reliability or unreliability. An unreliable person is psychologically incapable of following through with what they said they would do. It is not necessarily the case that an unreliable person was being dishonest when he agreed to do something he later failed to do.

It appears then that there are primarily normative and primarily descriptive senses of character, lack of character, and traits of character. In addition, there appears to be both positive and negative descriptive senses of traits of character. The various senses of these character terms can cause conceptual confusion. For example, take the case of the September 11 highjackers. It is unclear how to properly describe the highjackers. Following the attacks the general public bristled at the suggestion that the highjackers might be properly described as having been courageous. However, *something* seems correct about describing the highjackers as courageous. It appears so in part because it seems inappropriate to describe the highjackers as either cowardly or rash, the two typical vices associated with courage. The highjackers faced death for the sake of what they regarded as good, and thus cannot appropriately be described as cowardly. On the other hand, the highjackers fully recognized the dangers involved in their mission, and did not carelessly rush into danger ignorant of that danger. Therefore, describing them as rash is inappropriate. On the other hand,
the highjackers had a consistent and coherent deliberative framework which was stable and temporally extended. The fact they were committed to an object they regarded as good (whatever that object was) justifies the use of the descriptive sense of ‘character’ in describing the highjackers. However, the ambiguity in the use of ‘character’ allows some to describe the highjackers as courageous and others to reject such a description. What is at issue is the normative sense of ‘character.’ Those who describe the highjackers as courageous are appealing to the descriptive sense of the character term ‘courageous.’ Those who refuse to ascribe courage to the highjackers are appealing to the normative sense of the term. ‘Courage’ cannot be used to describe an agent whose goals, ends, and commitments are evil.

What I have outlined here suggests a way of thinking about the role of character in virtue. What explains the difference between a courageous agent and the highjackers is that the assessment of the courageous agent involves a positive affirmation that the courageous agent possesses character as well as a positive normative assessment of the objects of this agent’s commitments. The highjackers, on the other hand, receive a positive affirmation of the possession of character but a negative normative assessment of the object of their commitments. Virtue, then, always requires a positive-positive assessment of an agent’s character. Therefore, the possibility of being virtuous requires the positive fulfillment of both a normative condition and a psychological condition. In order for virtue theory to be a viable option, it must be possible for human beings to have as the objects of their commitments things of genuine value. It must also be possible for human beings to have character.
Having character is a condition of possibility for being virtuous. Another condition of possibility for being virtuous is that the agent’s commitments as embodied in his or her self-governing policies have as their object the correct conception of the good. The possibility of being virtuous is the result of having character and having a correct conception of the good.\footnote{A complete account of a moral psychology of character would require a descriptive account of vice, incontinence, continence, beastliness and wantonness. I have omitted such an account in order to concentrate on an account of the role having character plays in being virtuous.}

4.4 Having Character and Being Virtuous

At this point in my argument one might ask the following questions. Why is having character a condition of possibility for being virtuous? How is it that the temporally extended planning theory of human agency and what you are calling the descriptive sense of character are connected in a meaningful way to the possibility of being virtuous? What you have done so far is provide a descriptive account of the temporally extended planning theory of human agency together with a normative account of agent rationality with respect to the dependence of human beings on standards of consistency and coherence as described in Bratman’s historical principles of rationality, while rejecting a foundational account of agent rationality based on desire satisfaction. You then defined character in terms of the accounts you have presented. But why should we follow you in thinking that the account of “character” which you have given bears any explanatory relationship to the possibility of being virtuous?
My answer to this challenge is to argue that the virtues are understood in large part by reference to the functional roles they (are believed to) play in human psychology and the moral life. If I can show that the account of having character I have provided in the terms of 1) a descriptive account of the temporally extended planning theory of human agency, 2) a normative account of bounded practical rationality and 3) a good-based foundational account of reasons for action, can fulfill the functional roles the virtues (are believed to) play, then I can draw the connection between having character and being virtuous.

Recall from Chapter 1 that the functional roles the virtues (are believed to) play are: 1) that traits of character operate in human agency; 2) that traits of character be relatively stable and entrenched; 3) that the stability of entrenched traits of character contribute to an agent’s development of self by maintaining moral harmony; 4) that the voluntary acquisition and modification of traits of character contribute to an agent’s development of self by maintaining narrative unity; 5) that traits of character are constitutive of eudaimonia, of flourishing or living well; 6) that traits of character be subject to voluntary acquisition and modification; 7) that the voluntary acquisition and modification of traits of character allow for the possibility that an agent is morally responsible for the character he or she constructs for him or herself; 8) that traits of character allow an agent consistently to perceive objectively appropriate reasons for action as appropriate reasons for action; 9) that traits of character properly guide action on the basis of objectively appropriate reasons for action; and 10) that traits of character properly motivate an agent to act for, or because of, objectively appropriate reasons for action.
The synthesis of the descriptive account of the temporally extended planning theory of human agency, the normative account of bounded practical reason, and a good-based account of reasons for action together provides a straightforward way of accounting for these functional roles. I have summarized these resources in Table 4.1. I will say more about each point in what follows.
### TABLE 4.1: RESOURCES AVAILABLE FOR MEETING ANSCOMBE’S DEMAND

<table>
<thead>
<tr>
<th>FUNCTIONAL ROLES OF THE VIRTUES</th>
<th>RESOURCES OF PLANNING THEORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Operate in <em>Human Agency</em></td>
<td><strong>The Temporally Extended Planning Theory of Human Agency</strong> is based on premises drawn from observing and experiencing human agency. Human beings are rationally bounded, planning agents who are temporally extended and reflective beings.</td>
</tr>
<tr>
<td>2. Stability and Entrenchment</td>
<td><strong>The Inertia of Intentions, Plans and Policies</strong> arising from reasoning-centered commitment involves a disposition on the part of an agent to retain an intention, plan or policy without reconsideration except when an agent is confronted with a problematic situation or a moral crisis.</td>
</tr>
<tr>
<td>3. Moral Harmony</td>
<td><strong>Self-governing Policies</strong>, by embodying the fundamental commitments of an agent, which act as foundational framework reasons, facilitate the formation of lower order reasons for action which are consistent and coherent with respect to the agent’s deliberative framework. Except in cases of moral crisis, an agent with self-governing policies cannot be committed to one conception of what is right and yet have motives which conflict with this conception.</td>
</tr>
<tr>
<td>FUNCTIONAL ROLES OF THE VIRTUES</td>
<td>RESOURCES OF PLANNING THEORY</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>4. Narrative Unity</td>
<td><strong>Commitment to Justifying Ends</strong> allows an agent to develop the ability to deliberate in ways that maintain one’s integrity, the ability to assess one’s successes and failures over the course of time, and the ability to recognize reconsiderations of prior commitments as arising from problematic situations generated by the one’s prior commitments.</td>
</tr>
<tr>
<td>5. Constitute <em>Eudaimonia</em></td>
<td><strong>The Facilitation of Intra- and Inter-Personal Coordination</strong> provided by the proper operation of intentions, plans, and policies allows a human being to flourish, live well, or achieve <em>eudaimonia</em> by allowing human beings to reason in characteristically human ways, construct a life of sufficient unity, and sustain the social relationships in which <em>eudaimonia</em> is sought.</td>
</tr>
<tr>
<td>6. Voluntary Acquisition and Modification</td>
<td><strong>Multi-tiered Practical Reasoning</strong> allows an adult agent, when confronted with a problematic situation, to reconsider his or her intentions, plans and policies that together constitute his or her deliberative framework.</td>
</tr>
<tr>
<td>FUNCTIONAL ROLES OF THE VIRTUES</td>
<td>RESOURCES OF PLANNING THEORY</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>7. Moral Responsibility</td>
<td><strong>Commitments Embedded in Multi-tiered Reasoning</strong> – Given that an agent’s intentions, plans, policies, and commitments to justifying ends are either voluntarily acquired or open to voluntary modification through reconsideration in the face of problematic situations, an agent is ultimately responsible for the shape of his or her deliberative framework, and thus responsible for his or her character.</td>
</tr>
</tbody>
</table>
| 8. Moral Perception              | **Focusing Perception (positive)** - Prior intentions, plans, policies, and the commitments to justifying ends embodied in self-governing policies focus the attention of rationally bounded agents on the salient features of their circumstances amid the situational 'noise' by providing standards of relevance.  
**Silencing Perception (negative)** - Prior intentions, plans, policies, and the commitments to justifying ends embodied in self-governing policies provide filters of admissibility, allowing certain possible reasons for action to be silenced as they violate some normative condition of bounded practical reasoning. |
| 9. Guide Action                  | **Framework Reasons and Justifying Ends** guide action by restricting the range of available and acceptable options while also steering action toward proximate ends appropriately related to the justifying ends to which the agent is committed. |
### TABLE 4.1: RESOURCES AVAILABLE FOR MEETING ANSCOMBE’S DEMAND

<table>
<thead>
<tr>
<th>FUNCTIONAL ROLES OF THE VIRTUES</th>
<th>RESOURCES OF PLANNING THEORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Moral Motivation</td>
<td>Volitional Commitment</td>
</tr>
<tr>
<td></td>
<td>characteristic of intentions, plans and policies is a result of a conduct-controlling pro-attitude toward the object of an agent’s intention, plan or policy.</td>
</tr>
</tbody>
</table>
4.5 Fulfilling the Functional Roles of the Virtues

These functional roles are drawn from the contemporary philosophical literature concerning the virtues. Each is identified in an objection to a rival moral theory, such as consequentialism or Kantianism, in a reply to an objection raised against virtue theory, or in a positive account of virtue. I have provided no arguments to support the claim that each of these functional roles is necessary for the moral psychology of a virtuous agent. Nor have I provided any arguments supporting the claim that the list of functional roles I have presented here is exhaustive. Instead, my methodology has been to look at what contemporary authors are saying about virtue, and then to find a way of understanding how what they are assuming is possible. In this section of Chapter 4 I will be discussing how the descriptive account of the temporally extended planning theory of human agency, the normative account of the standards governing bounded practical rationality, and the good-based foundational account of reasons for action can together account for the possibility of being virtuous by providing an explanation of how character fulfills these functional roles. I will address each functional role, though I will be concentrating on a limited number of them.

4.5.1 Operate in Human Agency

In Chapter 1 I mentioned that part of the problem with ethical views that understand ethics as a set of rules is that these ethical views, such as Kantianism, have an unrealistic conception of the capacities of human rationality. Virtue theory, by contrast, is committed to providing a rich conception of the scope of the ethical
while also recognizing the true nature of human agency. For example, Bernard Williams argues that Kantian accounts of practical reason require equating practical reflection with detachment from personal interests and desires. Human agency is conceived of as unencumbered and unbounded. However, since human reflection and deliberation is essentially first personal and thus of necessity influenced by the social and deliberative frameworks within which an agent is operating, Williams concludes that Kantian practical reason is impossible for human beings. An explanation for why Williams is correct and why Kantian practical reasoning is impossible for us is that we are subject to bounded psychological capacities as described in the temporally extended planning theory of human agency.  

Charles Taylor and many others have argued that moral theories such as consequentialism reduce all human practical reasoning to mere instrumental reasoning. However, we have seen that the normative account of the standards governing bounded practical reason prohibit the reduction of human practical reasoning to mere instrumental reasoning. While on a desire satisfaction foundational account of reasons for action, the ultimate external assessment of agent rationality is made in terms of instrumental reasoning, assessments of plan-constrained practical reasoning cannot be made merely in terms of instrumental reasoning. In other words, even a desire satisfaction foundational account of reasons for action cannot reduce all

223 See Bernard Williams, Ethics and the Limits of Philosophy. (Cambridge: Harvard University Press, 1985). See Christine M. Korsgaard, “Skepticism about Practical Reason” in Creating the Kingdom of Ends (New York: Cambridge UP, 1996), 311-334, for a reply to Williams’s argument in which she argues that Williams’s characterization is based on the assumption that there are no unconditioned principles of reason.

human practical reasoning to mere instrumental reasoning. Similarly, a good-based foundational account of reasons for action also does not reduce all human practical reasoning to mere instrumental reasoning. However, according to a good-based account, even the external assessment of agent rationality is not made in terms of mere instrumental reasoning. Therefore, in both cases, the temporally extended planning theory of human agency recognizes not only the limitations of bounded practical rationality, but also recognizes the diversity and richness of human reflection and deliberation.

A viable virtue theory would be one that provides an account of practical reasoning that is consistent with the kind of rational beings we are; in particular, that is consistent with both our limitations and our abilities. We have just seen how the descriptive account of the temporally extended planning theory is consistent with our limitations, and how the normative account of the standards governing bounded rationality is consistent with our abilities. In addition, if what I have argued in Chapter 2 is correct, a viable virtue theory requires a consistent account of how good-based reasons for action can operate in human agency in a manner consistent with human limitations and abilities. These reasons for action enter into human deliberation by being embodied in self-governing policies which in turn provide foundational framework reasons. These foundational framework reasons provide filters of relevance and filters of admissibility with which an agent can screen the appropriateness of further possible reasons for action. Therefore, another way in which character operates in human agency is by allowing foundational framework reasons for action to operate as inputs into further practical reasoning. In this
understanding of character, character has a direct role in human agency and is not merely the result of agency which meets some set of externally defined criteria, such as a tendency to act in a particular sort of way. Instead, character is integral to the perception, deliberation, and motivation of the virtuous agent.

4.5.2 Stability and Entrenchment

In order to fulfill the functional role of operating in human agency, the psychological apparatus which is to account for the possibility of being virtuous must be stable and entrenched. Virtue theories often describe virtue as a habit, a disposition, or a state of character. To refer to virtue as a habit, a disposition, or a state of character is to indicate that part of what we mean by ‘virtue’ is that it is something stable and entrenched. The use of these terms with reference to virtue indicates that we understand virtue to be something which endures over time. ‘Habit,’ ‘disposition’ and ‘state’ cannot be used appropriately in contexts in which that to which they refer exists or endures momentarily. These terms, utilized by virtue theorists, suggest that momentary feelings, thoughts, and actions are insufficient to characterize an agent’s psychological apparatus as that of virtue. In addition, one cannot be said to have a habit, a disposition, or a state if he or she only exhibits the corresponding property intermittently. ‘Habit,’ ‘disposition’ and ‘state,’ at least in the contexts in which they are used by virtue theorists, all imply that they have a significant controlling influence over an agent, over both her deliberations and her actions, for an extended and uninterrupted length of time. Someone who possess a habit, a disposition, or a state does not succumb to temptation, temporary preference changes, temporal discounting, or social pressure. Therefore, as a result of being a
habit, a disposition, or a state, a central feature of virtue is its role in allowing an agent to remain temporally consistent.

A central feature of the descriptive account of the temporally extended planning theory of human agency and especially the normative account of the standards governing bounded practical reasoning is that it provides a means of accounting for temporal consistency despite the limits of bounded practical rationality. Recall that bounded practical rationality is the result of the facts that our most common practical decisions usually concern discreet, specific matters that do not range over our entire lives, that our decision making process cannot include exhaustive probabilistic characterizations of all the possible alternative futures, that our decision making process usually involves a type of selective attention in which certain facts about our selves and our lives are ignored, and that each of these facts contributes to the temporal pressures that bear on human deliberation. Recall that a central concern which the temporally extended planning theory of human agency was intended to address was the problem of limited agents acting despite an inability to deliberate concerning every possible course of action. Human beings compensate for these limitations by forming plans. Planning allows an agent to give his or her attention to certain extremely important variables in a situation while ignoring the vast number of possible influences on his or her decisions. Settling now on a plan enables present deliberation and reflection to influence an agent’s future conduct, thus minimizing the temporal pressure brought to bear on an agent at the moment of action. Central to the ability to effectively plan is the inertia characteristic of the reasoning-centered commitment attached to intentions, plans and policies. Inertia is a
disposition on the part of an agent to retain a prior intention without reconsideration, except when faced with a problematic situation. Inertia lends stability to an agent’s intentions, plans and policies. Since character is the result of relationships between an agent’s intentions, plans and policies, character inherits the stability embodied in an agent’s deliberative framework.

4.5.3 Maintain Moral Harmony

The stability of an agent’s deliberative framework also contributes to an agent’s moral harmony. Moral disharmony is the result of an unacceptable fissure between one’s motives and one’s judgments concerning right and wrong. Stocker argues that a mark of the good life is a harmony in both directions between one’s motives and one’s reasons, values, and justifications. Stocker’s argument is intended to show that modern moral theories (such as hedonistic egoism, classical utilitarianism, Moore’s utilitarianism, as well as rule-utilitarianism and deontology) all make moral harmony impossible, particularly with respect to personal relationships involving love or friendship. It is essential to the concepts of love and friendship that one care for the beloved and act for the beloved’s sake. However, the egoist believes his own pleasure to be the sole justification of moral action. The pleasures internal to a relationship of friendship can only be attained if one acts for the sake of the beloved. Therefore, the egoist, in seeking the pleasure internal to personal relationships, cannot embody his reason for action (seeking his own pleasure) in the requisite motive (acting for the other’s sake). The egoist must either

225 See Stocker, “The Schizophrenia of Modern Ethical Theories.”

173
abandon the egoist reasons for action, or abandon the motive necessary to receive the pleasures internal to a loving relationship. Stocker makes similar arguments against classical utilitarianism and Moore’s utilitarianism. In each case, since the agent sees the other as a source of value to be maximized, he cannot act from a motive capable of maximizing that value (acting for the sake of the other) without abandoning his reasons for action (to maximize pleasure or goodness). Finally, Stocker argues that Kantianism is no better off with respect to the problem of moral harmony. An agent can attempt to bring his or her motives and reasons into line with one another by attempting to embody Kantian reasons in his or her motives. The Kantian approach requires an agent’s motive to be one of acting from respect for the moral law. Unfortunately, this procedure results in an agent who alienates him or herself from the value and goods internal to personal relationships involving love and friendship. This is the case because such an agent comes to act for the sake of duty or out of obligation rather than for the sake of the beloved. The most troubling aspect of the problem of moral disharmony is that on the dominate moral theories available to modernity, moral action becomes self-defeating (as the cases of the egoist and the utilitarian show), or else it becomes self-stultifying such that through the performance of moral action an agent is deprived of important sources of value, such as personal relationships (as the case of the Kantian shows).

The problem of moral harmony is one of embodying our reasons, values and justifications in our motives in such a way as to make it possible to preserve important sources of value. In other words, the agent who according to the theory is a moral exemplar must be able to act for the sake of all those things which are of
genuine value. Unlike Kantianism and consequentialism, which appear to promote either moral disharmony or an ethically impoverished life, virtue theory has the resources to account for moral harmony in a virtuous agent. Or so some virtue theorists claim. The assumption made by these virtue theorists is that the virtuous agent is motivated by what she recognizes to be of genuine value. So, for example, since the virtuous agent recognizes that certain forms of social relationships are constitutive parts of flourishing or living a human life well, she has reason to act for the sake of fostering those relationships. And since maintaining those relationships requires, as a constitutive means, that one act for the sake of the other, the virtuous agent has reason to act for the sake of the other. In addition, since the virtuous agent recognizes acting for the sake of the other as being of genuine value, takes acting for the sake of the other as a foundational reason for action, and has embodied this foundational reason for action in her self-governing policies, it follows from the nature of the commitments involved in self-governing policies that the virtuous agent is motivated to act for the sake of the other. The temporally extended planning theory of human agency recognizes that for creature 8 (the creature representing the psychological apparatus of the virtuous agent) a judgment that an end is a justifying end produces the motivation to perform the corresponding action. In this way, an agent’s reasons for action are embodied in his or her motives. Therefore, if one’s intentions, plans and policies embody a commitment to a particular other, then acting in ways directed toward the good of the other becomes the agent’s motive for action. Again, the problem of moral harmony is the result of a moral psychology that allows, indeed requires, an agent to create a disharmony between his reasons and his motives.
The solution offered by the moral psychology of character presented here is that the forms of commitment characteristic of the intentions, plans and policies of the virtuous agent guarantee that the virtuous agent’s reasons for action are in harmony with his or her motives.

4.5.4 Narrative Unity

A premise of Stocker’s argument concerning moral harmony is that a good human life requires some level of moral harmony. A human being, in order to have a good life, must be moved by what he or she values, and value that by which he or she is motivated. In addition, virtue theorists often argue that a good human life must be one which contains a narrative unity which unifies and makes intelligible the life of an agent. The observation that a good human life requires this kind of unity usually stems from noticing the tendency of other moral theories to regard the lives of human agents as divided into discreet temporal parts, particularly with respect to deliberation and choice. The temporally extended planning theory of human agency implies that this conception of human choice is flawed. Instead, an agent’s choices only make sense, and are only rationally justified, in terms of his or her prior deliberations. The narrative of the agent, the story that can be told about an agent’s prior deliberations, choices, and actions, is captured in the current intentions, plans and policies of an agent. In addition, the narrative unity of a human life has a teleological character. There is a sense in which a lived narrative has a trajectory. The trajectory of a life is the result of the operation of the objects of an agent’s self-

governing policies as foundational framework reasons for action. However, since these foundational framework reasons for action are open to reconsideration given the appropriate circumstances, the narrative which unifies a human life is always open-ended. Given that the non-foundational reasons for action are more easily reconsidered, there is always a great deal of play in one’s commitments. Despite the possibility of problematic situations and moral crises, the deliberative framework of a virtuous agent embodies the agent’s history of deliberation, decision, and choice as well as his or her commitments. It is by reference to this history that an agent is able to begin to make choices concerning what is best and good. Without a prior history and an envisioned future, there exist no criteria reference to which will serve as a means of making choices from among the nearly limitless range of possibilities available to an agent at any given moment. However, with access to her own deliberative history and envisioned future as captured in her intentions, plans, and policies, an agent is able to place her life in sufficient order to make intelligible choices.

Alasdair MacIntyre has argued that the virtues are necessary to sustain the narrative unity of an individual life in which an individual is capable of seeking his or her good.\textsuperscript{227} The account of character I am presenting provides an explanation of how the virtues can play this role in sustaining narrative unity. The temporally extended planning theory of human agency takes as a premise that human deliberation and human agency is temporally extended, and not merely into the past. Human agency, through the formation of future-directed intentions, plans and policies, is embedded in the deliberative framework of a virtuous agent.

\textsuperscript{227} MacIntyre, \textit{After Virtue}, Chapter 15.
policies, is dependent on a human agent’s ability to envision a future meeting some minimal level of specificity. By recognizing that past deliberation enters as inputs into present deliberation directed toward some future set of ends and goals, the temporally extended planning theory of human agency explains how the narrative unity of a human life is possible.

4.5.5 Constitute *Eudaimonia*

Michael Stocker argues that moral harmony is necessary for a good human life. Alasdair Macintyre argues that the narrative unity of a human life is a necessary condition of seeking the human good. Among virtue theorists, *eudaimonia* is a popular term to designate the goal or *telos* of a human life. Typically *eudaimonia* is translated from the Greek as ‘happiness.’ However, many virtue theorists reject this translation because of its tendency in modern English to refer merely to subjective feelings, typically feelings of pleasure or contentment. Increasingly virtue theorists are insisting that *eudaimonia*, if it is translated at all, be translated as living well, flourishing as a human being, or living a good human life. Therefore, it is appropriate to interpret Stocker’s and MacIntyre’s claims as asserting the role of moral harmony and narrative unity in *eudaimonia*. Many virtue theorists, most notably Alasdair MacIntyre and Philippa Foot, have argued that the notion of flourishing is a matter of fact concerning an individual’s development and proper utilization of certain capacities, abilities, and powers which he or she has in virtue of being a member of a particular species. So, to flourish as a human being may include, for example, the ability and capacity to think and deliberate well in specifically human ways, or the
ability and capacity to make progress toward the truth concerning what is good and best for us.

The account of character I have been developing has the ability to describe the various ways in which character helps constitute *eudaimonia*. One of these ways is to give an account of moral harmony. Another way is to provide an account of how a human life is integrated into a unified whole, both in the present and across time. In addition, character provides the psychological apparatus required for humans to deliberate well in distinctly human ways. Without character, humans fail to have the resources to fulfill their potential as independent practical reasoners. Without character, humans are deficient with respect to their ability to evaluate reasons because they do not have sufficient criteria in the form of framework reasons with which to evaluate possible reasons for action. Without character, human beings are also deficient with respect to their ability to have a sufficiently developed awareness of an imagined future.

Finally, future-directed intentions, plans, policies facilitate inter-personal coordination. In order for a human being to flourish, live well, or achieve *eudaimonia*, he or she must be able to sustain the social relationships in which *eudaimonia* is sought. A minimal amount of coordination is required to create an environment in which significant relationships can be sustained and flourish. It is within these relationships that the good life is sought through various forms of

228 Alasdair MacIntyre, *Dependent Rational Animals*, (Peru: Open Court Publishing Company, 1999)

229 These two deficiencies are pointed out as a result of Alasdair MacIntyre’s argument in *Dependent Rational Animals* concerning the transition from child to independent practical reasoner.
communal deliberation.\textsuperscript{230} Communal deliberation of the sort that can maintain the narrative unity of an agent often takes place within the context of practices. Sustaining practices across time also requires inter-personal coordination. Not only must the intentions, plans, and policies of the practitioners be sufficiently explicit, but they must be carried out in order for the practice to not break down as a result of lack of coordination. Therefore, the account of character I have been developing helps explain how the personal and social relationships and practices required for \textit{eudaimonia} are possible.

I have been arguing in the last few sub-sections that the account of character I have been developing allows for an explanation of how it is possible for an agent to be virtuous by explaining how character can fulfill the functional roles of the virtues. Now, the virtue theorist is committed to the claim that it is specifically virtue, not merely character, that must fulfill these functional roles. My argument, though, is that it is the psychological apparatus captured in my account of character that explains the possibility of being virtuous. The actuality of being virtuous would require an additional account of what foundational commitments are the proper commitments for a virtuous agent to have. However, as I have said before, this subsequent account is to be provided by an ethical theory, not a moral psychology. In other words, having character, while fulfilling these functional roles and necessary for being virtuous, is not sufficient for being virtuous. The possession of character merely fulfills the psychological conditions of possibility for being virtuous.

\textsuperscript{230} See MacIntyre, \textit{Dependent Rational Animals}.
4.5.6 Voluntary Acquisition and Modification

Sometimes it is objected that the emphasis which virtue theorists place on moral education and habituation renders the acquisition of virtue beyond the control of an agent. The possession of virtue becomes an accident of one’s personal and social circumstances. However, the account of character that I am developing has the resources to provide at least a partial response to this objection. Recall that the account of practical reasoning presented in the descriptive account of the temporally extended planning theory of human agency as well as in the normative account of the standards governing bounded practical reasoning emphasized the multi-tiered structure of practical reasoning. Without denying the influence of a proper moral education on the acquisition of virtue, one can readily see the multiple opportunities available to an agent for the reconsideration or retention of an intention, plan, or policy. While a child is dependent on his or her teachers in developing the abilities and capacities to evaluate reasons in light of his or her intentions, plans and self-governing policies, a child eventually develops sufficient independence to gain a large degree of autonomy with respect to the construction of his or her deliberative framework. Typically an agent is confronted with problematic situations in which he or she must evaluate his or her potential reasons for action in terms of his or her commitments as embodied in his or her intentions, plans and policies. Each occurrence of a problematic situation is a potential opportunity for reconsideration and rejection of some aspect of a agent’s deliberative framework.
4.5.7 Moral Responsibility

An agent is morally responsible for the state of his or her character as a result of the frequency of opportunities for revision available to him or her. Given that an agent’s intentions, plans, policies, and commitments to justifying ends are either voluntarily acquired or open to voluntary modification through reconsideration in the face of a problematic situation, an agent is also ultimately morally responsible for the content of his or her deliberative framework, and thus responsible for his or her possession of virtue. In particular, since an agent is capable of asking whether the nature of his or her commitments really are good and the best commitments to have, an agent is ultimately morally responsible for his or her possession of virtue.

4.5.8 Moral Perception

David Wiggins argues that a neo-Aristotelian theory of practical reason understands deliberation as a matter of practical perception of the relevant features of a situation that activates thought and reason regarding certain concerns and passions. However, the relevant concerns and passions cannot be embodied in algorithmic form because it is a feature of human agency that our concerns often make competing and inconsistent claims on us. When means-end reasoning begins, the practical deliberator always has the option of remaking, refining, reevaluating, or rejecting a concern or end. “The unfinished and indeterminate character of our ideals and value structure is constitutive of both human freedom and . . . of practical rationality.”

---

the practical syllogism the minor premise is specified by what strikes an agent as the most salient feature of the context in which he or she acts. A minor premise activates a concern that features in the major premise. The major premise is a concern that is not restricted or qualified in its application at a given time.

John McDowell draws on Wiggins’ account of practical deliberation when he asserts that virtue is knowledge. McDowell argues that a situation requiring kindness must be the kind agent’s reason for acting. In order for the agent to have the appropriate reason for action, he or she must be aware that the situation presents such a reason for action. A virtuous agent must have reliable awareness of the reasons for action present in a situation. Reliable awareness requires “a reliable sensitivity to a certain sort of requirement which a situation imposes on behavior.”

Reliable sensitivity is a perceptual capacity.

The role of certain perceptual capacities in virtue is central. I have not gone so far as to define virtue as merely “an ability to recognize requirements which situations impose on behavior.” However, I do think that recognition of the requirements that a situation imposes on behavior is the result of having virtue. This capacity is in part the result of having character. Prior intentions, plans, policies, and self-governing policies focus the attention of rationally bounded agents on the salient features of their circumstances amid the situational ‘noise’ by providing standards of relevance derived ultimately from what the agent values. In addition, prior intentions, plans, policies, and the commitments to justifying ends embodied in self-governing

---


policies provide filters of admissibility, allowing certain possible reasons for action to be silenced as they violate some normative condition of bounded practical rationality given the structure and content of an agent’s overall deliberative framework. Some possible options become unthinkable in the sense that to consider them would upset the stability of intentions, plans, personal policies, or self-governing policies already reflectively endorsed and not the subject of a problematic situation.

4.5.9 Action Guidance

The temporally extended planning theory of human agency provides an account of how, in creature 8, self-governing policies provide justified ends that shape and direct actions through deliberation and reflection. The action of the virtuous agent is guided by his or her future-directed intentions, plans, personal policies, and self-governing policies. As rationally bound agents, human beings must commit themselves now to actions in the future. Part of the reason that the problem of action guidance seems so pressing is that the critics of virtue theory, and even some virtue theorists, are committed to the standard model of decision theory that requires that all decisions be related to the choice of present options. I submit that it is a commitment to this unrealistic model of practical rationality, and not a commitment to the nature of morality, that drives some moral theorists to develop rules and algorithmic decision procedures for right conduct. Implicitly, these theorists recognize the bounded rationality of human beings. However, rather than recognize that bounded rationality is a feature of human nature and that we have developed a method to compensate, some moral theorists reduce the ethical such that
its content alone can guide action through the easy application of rules and principles to present options.

Hursthouse has made a related charge against the critics of virtue ethics. The critics of virtue ethics argue that virtue theorists not only have not, but cannot answer the questions ‘What ought I to do?’ and ‘What ought I not do?’ Therefore, virtue ethics is not a genuine alternative to the ethics of rules, but is at best a set of considerations that could supplement an ethics of rules. In her initial reply to this objection, she stresses that the objection depends on the assumption that a moral theory must codify morality; it must produce an algorithm for all moral choices. Hursthouse’s criticism of this foundational premise is that it is unjustified in a secular world. Without an appeal to a divinely rational creator God, a secular moral theory has no justification for the claim that all aspects of the moral world will fall within the realm of human rational understanding. Therefore, the belief in the complete codifiability of moral choice is unjustified.

The reply to the critics of virtue theory from the planning theory perspective is different from Hursthouse’s, but supplements it nicely. From the planning theory perspective, the continued appeal to rules is not merely an unjustified anachronistic holdover (as Anscombe, MacIntyre and Hursthouse would have us believe), but it is also the result of a particularly misguided attempt to overcome our bounded rationality by appealing to codifiable rules and principles and severely restricting the scope of morality to the realms of human life in which the rules and principles can be applied. It is as if the critics of virtue theory think that without an appeal to a simple rule or principle, human beings would be left with no other recourse but to be driven
to act on the basis of their strongest desires. On the planning theory of human agency, this dichotomy ignores the range of possible creature models that could capture our rational capacities. On their understanding, there exists only Creature 1 and some yet unspecified Creature 9, perhaps what Aristotle would call a divine agent.

Most adult human agents are planning agents and so at least comparable to Creature 4. The conduct of Creature 4 is guided by his or her plans, which resist reconsideration, control conduct, and provide input into practical reasoning. The fact that plans involve volitional commitment and act as inputs into practical reasoning allows them to guide conduct. Plans also provide standards of relevance and filters of admissibility for options considered in deliberation. These features of the planning theory further contribute to its ability to account for action guidance. However, we decide what plans and policies to have. What guides our decisions concerning what plans and policies to adopt such that these guide our actions?

The planning theory implies that we choose which plans and policies to adopt through a process of indentification. Identification is a decision to regard something as reason-giving and to regard it as an end toward which one’s actions are guided (given the coherence, consistency and stability of one’s overall intentional structure of future-directed intentions, plans and policies). One identifies with something when he or she takes it to be a justified end worthy of pursuit. One then constructs plans and policies directed toward those justified ends. So, action is guided through identification with a justifying end and the construction of plans and policies. However plans and policies are defeasible. Therefore, we need a way to guide our
reconsiderations of our plans and policies. The normative conditions of bounded rationality determine when plans and policies should be reconsidered. One such condition is that an agent regrets neither his reconsideration nor his non-reconsideration.

Bratman correctly notes that there must be a way to distinguish why it is appropriate in some cases to not abandon my prior intention or policy at the time of action while in other cases it is not appropriate to stick with my prior intention or policy at the time of action. The solution is not to appeal to some explicit rule which the agent must apply to the case, but rather to appeal to the nature of human agency. As temporally extended planning agents we view the future as a relevant concern. Temporally extended planning agents justify their actions to their future selves by maintaining the connection between their actions in the present and their (rough) conception of their end or good. Therefore, the anticipation of future regret helps us understand the appropriateness of the stability of an intention or policy. It is important to living out our commitments that we guide our actions such that our future selves at plan’s end will not regret what our present selves have done or failed to do at plan’s beginning.

One different version of the action-guiding objection is the so-called epistemological objection.\textsuperscript{234} If right action coincides with the actions of virtuous agents, one way to guide right action is to model our behavior on the actions of a virtuous agent. But how can we know who is virtuous? Robert Louden suggests two approaches. The externalist strategy attempts to observe behavior and infer the

character of the agent. This strategy fails because the virtue theorists are committed to the premise that there is merely a contingent connection between character and conduct. The internalist strategy is to identify virtuous moral psychology with some internal state. Louden suggests identifying virtuous moral psychology with neuro-physical states. Louden argues that this strategy fails because it assumes that we know who is virtuous in order to make the identification. The failure of virtue ethics to answer the epistemological objection raises skepticism concerning our ability to know who is virtuous and the value of our own motivations.

Louden is mistaken on at least two counts. First, more can be said on behalf of the internalist strategy. It is possible to interpret the strategy I have used as an internalist strategy. On my strategy, we identified, compiled and categorized the functional roles that the virtues would have to play in being virtuous. Then we looked for an account of mental states that could perform these functional roles. If we find these mental states in particular individuals, then we conclude that they have “the stuff of virtue” or the necessary mental apparatus from which virtues arise. We have a necessary condition for identifying the virtuous. This strategy provides us with the moral psychology of the virtuous agent. What it does not provide is the corresponding conception of the good that distinguishes the virtuous from the vicious. With this strategy, we can radically reduce the scope of our investigation into who is virtuous. In addition, we can ask these individuals what they take to be their justifying ends. Some justifying ends that may be honestly offered are clearly not

---

justified, such as pleasure, wealth, power, honor, etc. Those who offer these as justifying ends can be eliminated.

The second problem with Louden’s objection is that he argues that we must be skeptical of the value of our own motives if we cannot answer the epistemological objection. However, it is dangerous to rule out skepticism of one’s own motives. As creatures of bounded rationality, we must be aware of the fact that we can and do go wrong. This skepticism can be built into the functional roles of character without reducing their effectiveness. Indeed, if we follow Bratman, the kind of healthy skepticism that is necessary for rationally bounded agents to be virtuous is also necessary for character. The plans and self-governing policies which direct us toward our justifying end are stable but defeasible. For example, if my tradition were to undergo an epistemological crisis that required an adjustment to my conception of my justifying ends, then my character could not play its functional role if it did not allow for reassessment. Louden is, I think, overstating the moral capacities of a rationally bounded virtuous agent.

4.5.10 Moral Motivation

Given my reply to the problem of action guidance, the question as to why an agent should be motivated to perform any particular action becomes that much more pressing. However, the temporally extended planning theory of human agency provides an account of how justified ends motivate an agent with character to particular actions via the characteristic volitional commitment of intentions, plans, personal policies, and self-governing policies. It is important to recognize that the solution to the problem of action guidance and the problem of moral motivation are
exactly the same. The reason for this is that intentions, plans, personal policies, and self-governing policies act both as inputs into practical reasoning (and so can guide actions) and involve a commitment to certain actions and ends with which the agent identifies (and so can motivate action). These features of future-directed intentions distinguish them from desires or beliefs. Self-governing policies play a special role in moral motivation in that they allow an agent to take a stand with respect to his array of motivations. Recall that Bratman’s solution to the problem of subjective normative authority appealed to the idea that self-governing policies about which ends to treat as reasons for action provide the necessary connection between one’s mere pro-attitudes toward an end and one regarding that end as having normative authority in his or her motivationally effective deliberation. In order for this account to work, Bratman must provide an explanation of how agents develop self-governing policies regarding which ends to treat as reasons for action. Bratman’s solution is to argue that human agency also requires the concept of valuing.

An account of the moral psychology of character can only go this far with respect to the question of moral motivation. An account of the normative condition of possibility is required to prove an account of why any agent should value some set of ends over any other set of ends. This, I take it, is the work of ethics. Explicit in Anscombe’s Demand is the claim that the work of ethics can only begin after the work of moral psychology is complete. Therefore, I leave the remainder of the problem of moral motivation for a later date.
4.6 Conclusion

My positive argument in favor of a temporally extended planning theory of character is now complete. I have argued that the possession of character as defined above fulfills the psychological condition of possibility for being virtuous. The possession of character allows for the possibility that an agent possess the capacity consistently to recognize objectively appropriate reasons for action as well as the ability to act reliably for, or because of, those reasons. My argument began by compiling a list of functional roles that advocates of virtue theory claim the possession of virtue plays. I then presented an independently defensible account of human agency, the temporally extended planning theory of human agency. The argument of this chapter showed how the temporally extended planning theory of human agency can be extended to provide an account of character, and how this account of character can fulfill the functional roles identified by virtue theorists. By fulfilling these functional roles, the temporally extended planning theory of character meets the psychological conditions of possibility for being virtuous.

In the next chapter I move on to an examination of two possible objections to the account of character I have presented. One of these objections takes the form of direct challenges to my premises while the other takes the form of alternative accounts of character.
CHAPTER 5
INTENTION, COMMITMENT, AND COMMITTED INTENT

5.1 Introduction

I have now completed my positive argument in favor of a temporally extended planning theory of the moral psychology of character. The nature of this chapter is two fold. First, I will answer a series of objections to the temporally extended planning theory of character drawn from the work of Marcel Lieberman. Lieberman has argued that we cannot identify commitment and intention. If Lieberman’s objections withstand my replies, then the temporally extended planning theory of character is unsound. I will argue that Lieberman has offered no good reason to accept his contention that intention and commitment cannot be identified.

The second task of this chapter is to describe and critique a prominent contemporary alternative moral psychology of character offered by N. J. H. Dent. I will argue that Dent’s account of the moral psychology of the virtues is incomplete because he appeals to the unanalyzed concept of “committed intent” in order to provide for the regulation of what he calls sense-desire and passional-desire. However, the temporally extended planning theory of character offers a ready account of committed intent. In addition, the temporally extended planning theory of character makes explicit the temporally extended nature of human agency, a fact that Dent sometimes overlooks. Finally, the temporally extended planning theory of
character provides a clearer understanding of the way in which agents of character regulate their sense-desire and passional-desire than the explanation offered by Dent.

5.2 Intentions and Commitments

In Chapter 3 I discussed the nature of commitment. In that discussion I cited Marcel Lieberman’s discussion of the three central features of commitment, which are:

1. its stability over time and its capacity to be revised and reconsidered
2. its action-guiding force
3. its relation to self-understanding and identity

I added to this list a fourth feature:

4. its practical necessity

Lieberman noticed the similarities between his understanding of commitment and Michael Bratman’s account of future-directed intentions. The similarities which Lieberman notes are:236

a. Both are stable while allowing for revision, defeasibility, and flexibility
b. Both are conduct-controlling
c. Both can prompt and terminate practical reasoning
d. Both impose consistency and coherence constraints; they both serve as filters of admissible on options and require means-end reasoning
e. Both play an important role in intra- and inter-personal coordination

Despite these similarities, Lieberman resists the identification of future-directed intentions and commitments. To support his position, Lieberman cites two potentially significant differences between intentions and commitments. The first potential difference is that future-directed intentions are constrained by rational beliefs concerning what sorts of actions are under the agent’s possible control. One cannot rationally intend something over which one has no possible control, such as having good weather for a party. What is initially odd about Lieberman’s attempt to distinguish intentions and commitments in this way is that he appears to assume that being committed has no similar belief condition. However, this assumption is obviously false. One can no more be committed to good weather than one can intend good weather. What is required in both cases is at least a rational belief that the agent has a possible influence over the action or state of affairs in question. It appears, then, that both intentions and commitment are subject to rational belief constraints of this sort.

However, it is still an open question whether some other content of the applicable rational belief constraints might not distinguish intentions from commitments. Consider, then, the case of world peace. One might argue, as Lieberman does, that one can be committed to world peace while being unable to intend world peace. The idea appears to be that one can be committed to something while believing that one will have no influence over it. By contrast, it is not possible to intend something while believing that one will have no influence over it. In this way we can say, according to Lieberman, that anyone may be committed to world peace while we cannot say that any given individual may intend world peace. Only
those individuals who (rationally) believe that they will have an influence over global policies can be said to intend world peace. Yet, there appears to be no justification for this distinction on the temporally extended planning theory of human agency. If an agent is committed to world peace, presumably he or she will take actions to promote it and will refrain from actions which are inconsistent with such a commitment. The claim that is supposed to distinguish intentions from commitments is that while it is irrational to intend world peace if one believes that one will have little or no influence on world peace, it is acceptable to be committed to world peace despite one’s beliefs that one will have little or no influence in bringing it about. However, as we saw in Chapter 3, there is nothing irrational about intending to X even when one fully expects to fail to X. So long as an agent has a plan directed toward X, he or she can rationally intend to X despite his or her belief that he or she will fail (provided, of course, that the agent has at least a possible influence over global politics). Therefore, it appears that Lieberman fails in his first two attempts to distinguish intentions and commitments.

Lieberman’s second potential source of a distinction between commitment and intention is that certain sorts of commitments do not have any corresponding intentions. Certain sorts of commitments, such as commitments to friendship, do not provide a basis for forming a plan, and thus do not provide a basis for forming corresponding intentions. However, the temporally extended planning theory of human agency recognizes that human beings are not limited to the mere capacities to form intentions and plans. Rationally bounded creatures also form various kinds of policies which regulate, both positively and negatively, the formation of further
intentions and plans. So while it may be true that an agent cannot have an intention to friendship, it is false that an agent does not have intentions corresponding to friendship. In any personal relationship there exist policies the adoption of which are constitutive of an agent’s commitment to the relationship, policies such as seeking the beloved’s good for the beloved’s own sake. According to the temporally extended planning theory of human agency, these policies are general intentions which operate as inputs into bounded practical reasoning and govern the formation of lower-level intentions regarding one’s conduct with respect to, for example, one’s friendships.

Lieberman wants to deny that commitment can be cast as a kind of policy. According to Lieberman, the distinction between policies and commitments lies in how each is formed and, most importantly for our purposes, revised. Drawing on Bratman’s essay “Intentions and Personal Policies,” and David Velleman’s Practical Reflection, Lieberman distinguish 4 kinds of policies:

A. Unambiguous policies
B. Action ambiguous policies
C. Situation ambiguous policies
D. Dually ambiguous policies

An unambiguous policy is a policy to perform a particular kind of action in a particular kind of situation (e.g., always send a thank-you note after receiving a gift). An action ambiguous policy is a policy usually prohibiting a particular kind of action in a given situation while failing to provide a specification of the best positive action to be taken (e.g., never negotiate with terrorists). A situation ambiguous policy is a policy to perform a particular kind of action while leaving open the occasions on

---

237 Lieberman, Commitment, Value and Moral Realism, 74-75.
which one is to perform the action (e.g., give to charitable organization). A dually ambiguous policy is, according to Lieberman, a policy which specifies neither an action nor a situation calling for action (e.g., act appropriately with respect to friendship). Lieberman then asks whether commitment might not be best understood as a dually ambiguous policy. While he does think that our commitments are dually ambiguous, he does not think that it is appropriate to identify our commitments with policies.

According the Lieberman, policies have their natural application in discussions of institutions. In this, their natural context, policies have an impersonal nature. However, commitments are not impersonal in this way. Assume that commitment to a friendship is best understood as a dually ambiguous policy. Assume further, with Lieberman, that a policy is impersonal in nature. Then, “imagine explaining to a friend why you performed a certain favor for them in terms of your [impersonal] policy to do such things.” To cite a policy of this kind in an explanation of one’s actions toward a friend fails, according to Lieberman, to capture the personal nature of a commitment to friendship. Unfortunately, Lieberman’s argument depends on a false premise concerning the natural context of policies.

According to the temporally extended planning theory of human agency, policies are a natural feature of human agency. The formation and utilization of policies in bounded practical reason is the result of natural human capacities and abilities to compensate for their limitations. There is no necessary connection, nor even a natural connection, between a policy and an impersonal content. Indeed,

---

Lieberman cites no reason to believe that the content of a policy has to have any particular kind of content other than to note that institutions use impersonal policies. Perhaps it is true that impersonal policies are appropriate in institutional contexts. However, this does not show that impersonality is appropriate in other contexts. Indeed, in the case of friendship, what appears odd about Lieberman’s example is the content of the policy cited, not the mere fact that a policy is cited. Since it is a premise of Lieberman’s argument that the policy cited will be impersonal, the policy will have a form like “whenever someone, whoever he or she is, asks me to perform them a favor, I will perform it.” If I explain to my friend my performance of a favor, my friend should be taken aback, perhaps even insulted, that he or she has no special status in my practical deliberations. If all my actions with respect to my so-called friend are of this impersonal nature, there are grounds for supposing that no friendship in fact exists between us. Since this conclusion is reached on the assumption that an agent’s commitment to a friendship is identical to his or her policies concerning friendship, it follows, according to Lieberman, that a commitment to friendship cannot be identical to a policy. However, if the policy had the form “be helpful and accommodating toward my friend, S,” then there does not appear to be anything odd about appealing to this policy to explain to S why I performed a favor for her. I could say to S, “Look, I did this favor for you because you are my friend, and part of being a friend is being helpful and accommodating. Therefore, I have a policy of helping you out when you ask me.” Granted, the locution might be odd, but there is nothing obviously unsatisfactory about such a reply. Replying in this way does not reveal that I am alienated from the goods internal to my friendship to S. All
I have done is made explicit, in the terms of the temporally extended planning theory of human agency, that my deliberative framework is structured in such a way that when S asks me for a favor, I have already decided that I will help her. Helping S has taken on a deliberative priority within my deliberative framework. In the absence of a problematic situation that will force me to reconsider my friendship with S, I will act in a helpful and accommodating manner with respect to her. It is perhaps helpful to recall that in the context of the temporally extended planning theory of human agency a policy is a mental state. Therefore, in citing a policy to explain my actions to S, I am telling her something about my deliberative framework and about what I give deliberative priority. It does not seem odd to me to tell a friend that she and our relationship has a high deliberative priority; in fact, it seems to me that friendship requires that this priority be given. Therefore, if dually ambiguous policies can be personal, Lieberman has yet to provide any reason for distinguishing commitments from intentions (or policies).

Lieberman does present a more fundamental objection to the identification of commitments and intentions (or policies). According to Lieberman, policies are grounded in commitments to principles, values, and norms. Policies then are expressions of commitments. Therefore, intentions and policies presuppose the existence of an agent’s commitments. In contrast to David Velleman’s account of the role of policies in practical reason, Lieberman argues that policies are not directly connected to an agent’s self-understanding, but rather the space between self-understanding and an agent’s policies is mediated by commitment to certain principles, values, and norms. That is, policies are formed “against the background of
certain commitments." Polices provide a way of disambiguating one of the parameters of one’s commitments in order to facilitate practical reasoning with respect to the principle, value or norm embodied in that commitment. Therefore, commitments, while dually ambiguous, cannot be policies. Rather commitments are those aspects of practical reasoning of which policies (and intentions) are specifications.

Another argument against the identification of intentions (and policies) with commitments stems from Lieberman’s contention that the relation between policy revision and self-understanding is limited while the relation between commitment revision and self-understanding is dramatic. In particular, according to Lieberman, the reasons for reconsidering one’s commitments are more limited than the reasons for reconsidering one’s policies. Lieberman cites the following reasons for reconsidering policies:

1) mistaken beliefs about the circumstances
2) circumstances are not what one expected
3) change of belief-desire reasons for acting
4) one’s values change

However, “commitments reflect a greater personal stake of the agent in what she is committed to than do policies” and thus commitments are more intimately integrated into an agent’s self-understanding. Therefore, the revision of one’s commitments has a dramatic effect on an agent’s self-understanding, while the revision of a policy has a limited effect on an agent’s self-understanding since policies are only remotely

---

239 Lieberman, *Commitment, Value and Moral Realism*, 77

240 Lieberman, *Commitment, Value and Moral Realism*, 78.
related to self-understanding. Given the tendency of a revision of one’s commitments to upset one’s sense of self and identity, the revision of commitments occur under a restricted set of conditions. According to Lieberman, commitments are not revised under conditions (2) or (3) while they are revised condition under (4) and often reconsidered under condition (1). The exclusion of (2) and (3) from the possible conditions for reconsideration or revision is the result of an agent’s commitments “relation to the agent’s self-understanding and self-conception, while the inflexibility of policies [in so far as they are inflexible] is not necessarily, or very seldom, so related.”241 Policies can be adopted and revised without the self-assessment and reflection required for the adoption and revision of a commitment. “The agent’s view of herself as a certain kind of person, as a person with a certain set of values, is separable from the policies she adopts to a degree that is not the case for commitments.”242 As an example of this kind of disconnect between an agent’s commitments and the sort of policies an agent may adopt, Lieberman cites the policy adopted by a kind-hearted professor to refuse to discuss the grades he gives on term papers. The professor’s self-understanding and identity is wrapped up in his commitment to being friendly with and kind toward his students. However, he also knows that he is subject to being taken advantage of by his students. He also knows from experience that the initial grade he gives his students’ work is always the grade the work deserves. Therefore, to ensure that his students receive the grades they deserve and to prevent his students from taking advantage of his kind-heartedness, he

241 Lieberman, Commitment, Value and Moral Realism, 81.
242 Lieberman, Commitment, Value and Moral Realism, 82
adopts a policy which in large part conflicts with his self-understanding. He has adopted an authoritarian and dogmatic policy which conflicts with his kind-hearted and accommodating self image. This is possible, Lieberman argues, because the adoption of his grading policy is not particularly closely related to the professor’s self-understanding. Insofar as a policy is closely related to an agent’s self-understanding, it is usually because of an underlying principle to which the agent is committed of which the policy is a manifestation.

Lieberman’s argument against the identification of commitment and intention (and policies) rests then on the fact that our policies are more readily revised, and that our policies are manifestations of principles to which we are committed. I will address the second claim first. According to Lieberman, policies are specifications of principles to which an agent is committed. These principles play a central role in an agent’s self-understanding in a way that policies do not. However, Lieberman’s examples of principles to which an agent, such as the kindly professor, can be committed have a form such as ‘treat students fairly’ and ‘provide students with the best possible education.’ At least these examples of principles to which an agent can be committed appear to be policies. Granted, they are dually ambiguous policies, but they have a form similar to a policy. And we saw earlier that there is reason to reject Lieberman’s claim that commitments cannot be policies. One might wonder, then, if it isn’t the dually ambiguous nature of an agent’s commitments that distinguish commitments from policies. However, the mere fact that a commitment is dually ambiguous is insufficient to distinguish it from a policy. Recall the case of friendship discussed above. In that case, the policy I cited to explain to my friend why I
performed a favor for her was dually ambiguous; it cited no particular action I should take nor any particular situations in which I should act. So perhaps the distinction between intention and commitment lies entirely on the role commitments play in an agent’s self-understanding.

However, this basis of the distinction is also problematic. Lieberman ties the specificity and lack of ambiguity of a policy to its remote relation to self-understanding. This move suggests that a commitment’s role in self-understanding is a result of its generality. That is, ‘treat students fairly’ is general and ambiguous in a way that ‘do not reconsider grades’ is not. If we assume that only general and ambiguous principles can play a role in self-understanding, then commitments and intentions can be distinguished on the basis of generality, because it is this feature that allows a principle to play a role in self-understanding. However, it appears that with the proper beliefs a more specific policy could be intimately related to an agent’s self-understanding. For example, if a professor is committed to the principle of treating student’s fairly and believes (falsely) that a commitment to treating student’s fairly requires not reconsidering grades, then this professor’s grading policy is central to his identity and understanding of himself as a teacher. If the professor believes that the only way of living out his commitments is to adopt certain policies, then these policies become central to his self-understanding. Under these conditions, if one’s commitments resist reconsideration under conditions (2) and (3), then these policies would also resist reconsideration under conditions (2) and (3). Therefore, either these very specific principles are commitments (and not policies), or policies are sometimes subject to the same limited range of reasons for revision. Even the most specific
principle might resist reconsideration under conditions (2) and (3) if the principle is appropriately related to one’s self-understanding. Therefore, Lieberman’s argument for the distinction between commitments and intentions rests wholly upon the role commitments play in self-understanding and self-identity.

In reply to this final version of Lieberman’s argument, the temporally extended planning theory must appeal to the multi-tiered nature of bounded practical reasoning and its corresponding conception of self-understanding. Self-understanding on this view is the result of an identification with the framework reasons of one’s deliberative framework. The deliberative framework of a temporally extended planning agent is made up of self-governing policies, personal policies, plans, intentions, beliefs, acceptances, desires (and emotions). Each of these elements is dynamically inter-related such that a change in any one of them has a potential influence on other elements of the framework. Elements lower than the element of the framework which is changed are potentially effected as the agent modifies his or her deliberative framework in order to maintain consistency and coherence. Elements higher than the element of the framework that is changed are potentially effected by the generation of problematic situations and moral crises. To understand oneself is to understand what sort of reasons are to count as the best reasons for action. A change of any element of an agent’s deliberative framework will change an agent’s understanding of what sorts of reasons are to count as the best reasons for action. Therefore, it appears that if the criterion for distinguishing intentions (and policies) from commitments is the role commitments play in self-understanding, then there is no way to distinguish the two.
However, I think that the claim Lieberman is attempting to make is that a commitment is what the temporally extended planning theory calls a self-governing policy. A commitment is in some way foundational, more stable than an intention or policy, more entrenched, more intimately related to self-understanding, less likely to be reconsidered, and less likely to be revised or abandoned. Each of these conditions of commitment is a difference in degree and does not necessitate a difference in kind between intentions (and policies) and commitments. The temporally extended planning theory of character which I have developed in this dissertation is not subject to the objections raised by Lieberman. There appear to be no reasons to insist on a distinction between intentions (and policies) and commitments.

There exists reason to suppose that intention and commitment can be identified. In the next section I will describe and critique an alternative to the moral psychology of character I have presented. We will see that this alternative depends on the operation of something called ‘committed intent.’ I will argue that the temporally extended planning theory of character is superior to this alternative moral psychology of character because the temporally extended planning theory can provide an account of committed intent while the alternative leaves the concept unanalyzed.

5.3 Committed Intent

N. J. H. Dent has provided a well-developed account of the moral psychology of the virtues in the book, *The Moral Psychology of the Virtues*. Dent’s account is valuable and I think it is important to examine his account in considerable detail. However, I will argue that Dent’s account of the moral psychology of the virtues is incomplete because he appeals to the unanalyzed concept of “committed intent” in...
order to provide for the regulation of sense-desire and passional-desire. However, without an account of “committed intent,” he has not adequately accounted for the moral psychology of the virtues. I will also argue that the temporally extended planning theory of character offers a ready account of committed intent. In addition, the temporally extended planning theory of character makes explicit the temporally extended nature of human agency, a fact about human agency that Dent recognizes but sometimes overlooks. Finally, the temporally extended planning theory of character provides a clearer understanding of the way in which agents possessing character regulate their sense-desire and passional-desire than does the explanation offered by Dent.

5.3.1 Dent’s Moral Psychology of the Virtues

Dent’s project investigates what traits of character consist in and how the actions of an agent can express a trait of character. He begins by examining in abstract three active powers of the human soul: sense-desire, passional-desire, and rational-desire. On his account, these three powers are central to determining the actions of a virtuous agent. He then looks at the connections and relationships between these three powers that give rise to the virtues. Dent is heavily influenced by Plato and Aristotle in giving an account of how some desires “are based on or answerable to practical reason for their existence, strength, and direction, and thus that our purposes and concerns are open to rationally reflective assessment and
modification.” An account of virtue will explain how this is possible by describing the connections and relationships between desires, passions, and their regulation.

According to Dent, the active powers of human beings play a central role in determining the actions of an agent. An active power is “a source of human activity, a capability for acting in certain ways in relation to certain things.” The three active powers are varieties of desire, which Dent describes as good-independent desires, somewhat good-dependent desires, and wholly good-dependent desires. To say that a desire is good-independent is to say that the agent does not need to make any evaluative judgment concerning his or her object of desire; the object of the desire is naturally desired. Good-dependent desires require the agent to make a prior evaluative assessment of the object as worthy of desire.

5.3.2 Sense-Desire

Dent describes sense-desire this way: “To experience a sense-desire is to experience an inclination to secure, for oneself, the enjoyment of some sense-pleasure, some pleasure which comes from, or reposes in, the gratification of one (or more) of the five senses.” Dent argues that coming to have a sense-desire does not require valuing having the experience, but is rather “a natural response of the human organism to having had such pleasurable experiences.” Since sense-desires arise,

---

at least initially, independently of valuing the pleasure and independently of any form of deliberation, sense-desires are good-independent.

Sense-desire produces an inclination in an agent to secure the object of pleasure, independently of deliberation about the nature of the end. This inclination is the first step of a natural process to secure the object of sense-desire, although this natural process can be interrupted at various stages. “It will be precisely by interrupting the continuum of unfolding [sense-] desire into action that we will be able to effect the regulation of our action upon sense-desire.”

247 This non-deliberative continuum toward securing an object of sense-desire is the “original structure,” or the basic case (or the primitive case) of human action.

5.3.3 Passional-Desire

It would be a mistake to reduce all human acts to actions derived from sense-desires. Therefore, Dent distinguishes passional-desire from sense-desire. Passional-desire does not arise directly from good-independent pleasures and pains (as sense-desires do), but requires, as a necessary condition, that the agent make some valuation (though the valuation may not be obvious to the subject). In this way, passional-desires are similar to passional responses. For example, anger does not result from perceived injustice unless the agent perceives the actions of another as harmful or damaging, and the object of the damage is valued by the agent. Therefore, a


248 Compare the primitive case of human action to Bratman’s lower creatures. Also compare to the similarities between animals and human children noted by Alasdair MacIntyre in Dependent Rational Animals.
passional response requires the agent to have made an evaluative judgment of some sort.

Dent argues that the good-dependence of passional responses carries over to the associated passional-desire. Because a passional-desire does not directly involve a value judgment, Dent describes passional-desire as somewhat good-dependent. Active-passional-desire differs from other passional-desires in that active passional-desires are those whose “onset typically includes the propensity to engage in directed activity [activity directed to the attainment of some goal].”  

The goal of the activity is determined by the evaluative judgments of the agent, but the desire to act does not result from deliberation. Dent describes the onset of activepassional desire this way:

Some feature of a person’s situation, or something that (as he supposes) has happened or will happen, is believed by him to be in some way and to some degree bad, undesirable…. This belief induces in him, not after deliberation, but as an immediate unpremeditated response, a desire to modify the situation in just such a way as to rectify it in the respect in which he believes it to be undesirable.

The occurrence of active passional-desire often requires as a necessary condition a negative evaluative judgment on the part of the subject concerning some feature of his present situation (though not necessarily concerning himself). This negative evaluative judgment is based on the belief that something about the situation is undesirable, and the passion is rendered active by prompting the agent to change his situation in appropriate ways. However, even in cases of negative evaluative


judgment such as this, Dent argues that the ultimate source of all active passional-desire is ‘love.’ By ‘love,’ he means, “the investing of concern or care in something as being something of importance, value or good to one in one way or another.”

He also says, “Love is, I suggest, a response to the perception of positive value, and underpins all our other emotional responses” and brings with it a concern for the good of the object for its own sake. Love then is not properly speaking an emotion or a passion, but the source of emotion and passion. Dent summarizes his account:

I am claiming, therefore, that the following structure exists in the passions. Basic to all of them is the establishing of a loving attachment to, a concern for the good of, something or other….and in such an attachment the subject desires the well-being of what he loves, in whatever shape or form that may take, and is moved to act to ensure that well-being. If and when what we wish well is well, whether through our effort or not, we ‘celebrate’ this in our feeling elation, joy, [or] delight at this. If and when what we wish well fares ill then we shall be apt to respond with anger, fear, envy, pity, etc.; which particular passion is involved will depend upon what is loved and how its well-being is threatened or suffers. By these passions we are moved to wish to remedy the ill, and to act to procure this remedy. If and when we are unable to procure the appropriate remedy, then we shall be apt to respond with the mourning passions of sorrow, despair, melancholy etc., wherein we mourn the damage to what is loved.

251 Dent, The Moral Psychology of the Virtues, 82-83.

252 Dent, The Moral Psychology of the Virtues, 82.

The passions’ dependence on evaluative judgments is the way in which the passions share in a rational principle and the way in which passions can come to be rational, correct or appropriate. Loving attachments also form our identity. “Our particular nature, as the individual being we are, is constituted by the abiding concerns which our course of life embodies and enacts; and these are provided by the objects of our abiding loving commitments.” Dent’s use of ‘commitment’ is crucial here, as is his use of ‘abiding.’ I will return to Dent’s use of these concepts below, arguing that he needs but does not provide a way to integrate the loving commitments into human agency.

Nonetheless, Dent describes the way in which the active passional-desires are induced or evoked in an agent.

In no case does the subject instigate the desire himself, but it comes upon him. Even if only slight interest in procuring a certain [general] end is engendered, this [general] end has not been decided upon as a desirable one by the subject. This is a good part of what is meant when it is said that we suffer or undergo passion…. By the onset of a passion, itself a way we are ‘acted upon,’ we are induced to desire certain things, a second way we are ‘acted upon.’ Certain things hold our interest and in a measure we are in thrall to them, under their influence and direction and we are not entirely self-directing, masters of the placing of our concern.

---

254 See Aristotle *Nicomachean Ethics* 1102b 13.


256 Dent, *The Moral Psychology of the Virtues*, 72. At this point Dent is only entitled to the desire to remedy the evil on one’s situation. The desire to any more particular end must be the result of means-end reasoning. That is why I include general in the quotation above.
Onset of active passional-desire not only modifies attention and prompts action, but also influences the thoughts of the agent, often in ways that are undesirable. Dent’s example is the way in which thoughts about the present situation can become obsessive and distorted in the presence of a passional-desire to remedy some evil. For example, one might come to think ‘Life is not worth living until this evil is remedied.’ In addition, invoked active passional-desire can cause us to overestimate the desirability of certain remedies or the extent of the evil present. Therefore, passional-desire must not rule our lives, but must be coupled with practical reason.

Dent argues that one task of practical reason is to prevent us from loving wrongly. There are two ways one can love wrongly. The first is to make an error concerning the object’s possession of certain qualities that it is in fact appropriate to love. The second way to love wrongly is to love something to excess, to attach more importance to something than it deserves, or to take it to be more valuable than it is. Dent describes the tasks of practical reason this way:

It is…the single most important aspect of the exercise of practical reason to make our love intelligent and appropriate, to make the nature and degree of our involved concern with various goods appropriate to their true weight and significance. Practical reason, duly employed, should enable us to form a realistic and stable sense of the relative importance of the variety of goods to which we are attached, and to regulate our concern for these to the degree to which it is merited….Practical reason takes up, orders, sifts, regulates, and extends the often fragmentary and incoherent perceptions of value and concerns for value we possess in our passions, and moulds these into an integrated whole which could enable a man’s life to
As I have argued previously, a failure to distinguish the psychological aspects of virtue from the normative aspects of virtue leads to conceptual confusions. In Dent’s case, he has failed to distinguish the psychological aspects of character from the normative aspects of practical reason. In the passage above, the highlighted phrases are descriptions of the psychological tasks of character while the other phrases are descriptions of the normative tasks of practical reason. The result of Dent’s failure to distinguish the normative and the psychological is to misconceive the true task of practical reason, which is to make true judgments concerning the good and the best. Character, along with practical reason, is a necessary condition for virtue. The task of character is to stabilize and regulate our concern for the good and order, sift, regulate, extend and mold the often fragmentary and incoherent perceptions of value gained from practical reason into an integrated concern for value that can enable a man’s life to have some overall significance and lucidity.

Dent argues that since passions simply “obey” the subject’s evaluative assessments, practical reason has a role in ruling the passions through its regulation of the agent’s assessments. Since we do not have an immediate and natural passional response to everything of value, practical reason “can, should and [does] take them [objects of perceived value] into account and incorporate[s] them into the direction of our lives by deliberate election to make them our concern and by resolving upon

acting for their sake.” 258 Since passions and their claims on action can conflict, practical reason must also be able to resolve these conflicts by making comparative assessments and directing action accordingly. Practical reason must be able to counteract the tendency to be set upon by conflicting and stronger passions or inclinations away from the action initially determined to be appropriate. Practical reason has the job of integrating the values of an agent into his deliberative processes such that the agent becomes an independent practical reasoner with a sufficient unity of self and life. Practical reason allows us to become “the authors of our own lives, not those who unknowingly carry out the wishes of others, enacting the influences that molded us.” 259

A major weakness of Dent’s account is that he fails to recognize the bounded rationality of human beings. He is placing an unrealistic burden on our active rational abilities. If the integration of the self and the resolution of internal conflict is the responsibility solely of practical reason, the virtuous agent will constantly be deliberating whether to allow his or her evaluative judgment to stand, or whether to reassess his or her prior judgment. There is no way for an agent to determine when and if this decision concerning the appropriateness of an action is decisive. 260 Therefore, the evaluative judgments determining an action to be appropriate will constantly be open to challenge, not from conflicting passions, but from practical reason itself. Relying on an active power to produce the stability characteristic of

258 Dent, The Moral Psychology of the Virtues, 90. This passage includes the same failure to separate the tasks of practical reason from the tasks of character that I mentioned above.


260 This is the same argument Bratman deployed against Frankfurt. See above.
virtue will never succeed. The underlying problem is that Dent’s account remains within the belief-desire model of reasons for action (though he clearly wants something else). Dent analyzes sense-desire and passional-desire by first abstracting them from actual human agency. The result is that core features of human agency are overlooked in the analysis, particularly temporal extension. The reason he does not give an account of the temporally extended agency of human beings is that the desire-belief model of human action reduces all human deliberation to deliberation concerning what to do now, at the moment of action.

5.3.4 Rational-Desire

Beliefs do not govern and direct our lives and action only via the evaluative judgments that give rise to passional desires.

We can and do form desires upon the basis of our judgments of good and evil, and we do not always need to have desire evoked from us if judgments are to direct our purposes and conduct. Such desires I shall call ‘rational desires’ in that they are desires, for example, to do something which the agent has upon the grounds that, for the reason that, it would be best that he do this thing, as he believes.261

One way to get at the nature of rational-desire is to consider rational deliberation - deliberation concerning conflicting desires in an attempt to determine which to allow to influence conduct.

According to Dent, deliberation involves 1) the consideration of the value or importance of the actions, A and B; 2) the emergence of a decisive preference for one

action, A; and 3) the formation by the agent of a resolution or intent to A because it is
the best thing for him to do in the circumstances.

In his initial account of deliberation, Dent has again omitted the temporally
extended nature of deliberation and human agency. In addition, the account of
deliberation is an account of deciding, with explicit reference to the good and the bad,
which desire is to directly influence conduct in the present. Deliberation is confined
to the desire-belief model and to comparative assessments. It does not allow for the
influence of prior intentions to be direct inputs into the formation of rational desire.
Dent’s account of deliberation requires “going all the way back.” That is, his account
requires that a deliberating agent, at the moment of action, explicitly consider how
each of his available options directly relates to what he takes to be good and best
given the circumstances. However, part of those circumstances is the limitations of
the agent, not merely in his particular rational abilities and his moral character, but his
limitations as a human being who is rationally bounded. Therefore, given the
practical constraints on practical reason, deliberation can rarely be done with direct
reference to the good and the bad.

Dent also argues that the virtuous agent has to be able, when asked, to refer
his action to a conception of the good, however limited and to whatever extent he is
able. However, Dent admits that the virtuous agent does not need to explicitly
consider his action’s relation to the good and the best. Therefore, we can argue that
Dent’s account of deliberation is either inaccurate or counterfactual. In either case,
he has not given a good account of the moral psychology of the virtues.
Dent is arguing that deliberation results in a deliberative desire (a decision), which is equated to a resolution or an intention to do something *now*. On Dent’s account, rational desire is equivalent to a present-directed intention. But, if all rational desires are present-directed intentions, how can the results of deliberation influence future deliberation, how can rational deliberation regulate future passional responses, and how can rational deliberation result in the stability of passion Dent requires for virtue?

The judgments of deliberation influence conduct, but it is not clear how, as Dent admits. “We need, that is, to show in what way taking something to be a reason for purposing or doing something is an action-directing state, a conative state.”

Dent presents his account of reasons for action in terms of means and constitutive means to achieving what a subject takes to be a satisfactory life. “My general concern is to understand the general character of the thought someone has when he thinks, possibly wrongly, of there being a reason for him to pursue this or to undertake that, to understand the nature and implications of such a thought.” The implication is that the agent must have discovered a decisive reason to undertake that course of action. “It is not until one has concluded that there is overall a decisive case, or as good a case as is needful or possible, for one particular course of action that one forms the resolution of embarking on it.” A decisive case is made by showing how the action is a means to or constitutive of the agent’s conceptions of the good life. Therefore, the ability to direct one’s conduct by reasons for action requires

---


1) a conception of the good life (of the good) and 2) a “concern to make it the case that the conduct of one’s life will materially realize this conception.” Dent puts it this way:

A rational desire involves a belief that something has importance to the conduct of our lives and the determination to give place to that in our activity…. It is a form of self-determination by which we embody in our conduct our conception of what it is appropriate that we should do…. To say that someone has a rational desire to do something is to say that, providing the reasons to do it make a sufficient case in his eyes, he will do it; it is to say that he will act in view of those reasons.

Dent goes to great lengths to distinguish rational desire from the desire associated with sense-desire and passional-desire, both of which involve a natural propensity to action. Rational desire involves nothing of the sort -- quoting Anscombe he says that rational desire does not involve the ‘prick of desire,’ a need that we feel needs to be satisfied.

5.3.5 Regulating Desire

It is possible for the various modes of concern involved in the various kinds of desire to become so interconnected so as to direct and guide the actions of a single self. Together, the various modes of concern can be integrated into a single yet complex mode of concern. “An account of the fully and properly integrated condition

---

of these concerns is an account of the structure of that psychological state which comprises a man’s possession and exercise of a virtue.”267

5.3.5.1 Regulating Sense-Desire

The virtue of temperance is an exception in that there cannot be a full integration of sense-desire and rational-desire. Since sense-desire in aroused in us by nature, practical reason can only really endorse the pre-existing desire -- endorse what we already want by either judging the want and the indulgence of the want as good or innocent, -- or by judging the indulgence as better than the alternative of frustration. Nonetheless, “[the virtuous agent] acts deliberately and intentionally for the sake of enjoying that pleasure, even though the proximate source of his conduct still remains his state of aroused [sense-] desire.”268 The sense-desires can be “harnessed” in support of living a suitable life. Dent is insisting that the harnessing or endorsing of sense-desire requires deliberative reflection at the moment of action. “Between the arousal of desire and its enactment we often do, often should and often can interpose reflection. And the possibility of doing so enables us to incorporate the purpose such desires incite us to undertake into a deliberatively understood and sought plan of life, and enables us to make the actions we undertake on desire embody our deliberate intentions.”269 But what if we cannot endorse the promptings of sense-desires? “What if anything can be done in such a case to make it that the frequency, intensity, and persistence of sense-desire become congruent to the judgment of practical reason

about the appropriate degree of importance to be placed upon the enjoyment of sense-pleasure? It seems to me that the basic mode of regulation and ordering of which sense-desire is susceptible is through ‘inhibition’ of its expression.\textsuperscript{270} Inhibition of the expression of sense-desire is achieved in part through detachment from the claims of sense-desire, deliberatively resolving to refrain from acting on the basis of sense-desire. This may prevent action, but not the moderation and “harnessing” of sense-desire. However, since just the thought of and anticipation of a pleasure is a partial exercise of sense-desire, and, as Aristotle observed, the exercise of sense-desire increases its intensity and frequency, it appears that inhibition of the expression of sense-desire may also require ‘expunging’ the thought and anticipation of the associated pleasure from one’s thoughts and concerns. “A desire which is no sooner felt than it is thus expunged might well be less and less likely to recur.”\textsuperscript{271} In addition, revulsion to action may diminish pleasure and so diminish desire. Refusal to identify with a “deviant desire” may also diminish desire. If one is successful in regulating the expression of one’s sense-desires through one or more of these means, over time a congruence will emerge between the intensity and frequency of one’s sense-desires and one’s judgments, such that one can come to endorse the sense-desires one continues to have. In this way one can come to develop the virtue of temperance.

\textsuperscript{270} Dent, \textit{The Moral Psychology of the Virtues}, 136.

\textsuperscript{271} Dent, \textit{The Moral Psychology of the Virtues}, 138.
5.3.5.2 Regulating Passional-Desire

Passional-desires are significantly different from sense-desires in that the principle of their occurrence depends on evaluative judgments after deliberation, and so they are more directly regulated by rational reflection. Only by changing an evaluative judgment can an agent directly modify his or her propensity to have a passional response with respect to the object of the evaluative judgment. The inhibition of passional-desire will not result in a change in the intensity, frequency, and persistence of the passional-desire. Since passional-desires are good dependent, Dent argues that inhibition will fail to alter our evaluative judgments and so fail to modify our passional-desires. Rationally ordered passional-desire is only exhibited in a human life through the endorsements of value.

5.3.6 Committed Intent

When explaining the ways in which an agent’s endorsements of value are expressed, Dent offers the following example.

We look for, in the virtue of kindness, a committed intent to answer the claim of the needs of others which goes beyond the occasional effect of quickened feeling. That committed intent will include seeking to make the liability to a passional response to the needs of others more than merely occasional. One who takes the claim of others to receive help when in need as a matter which merits concern will try to resist any alteration in or disappearance of his tendency to be moved by their plight. He will endeavor to suppress the growth of inclinations which may obstruct or remove his tendency to this responsiveness; he will so order his thoughts, expectations, hopes, and fears so as to procure and to secure a steady constancy of regard to the needs of others, a steady disposition to have his concern
awakened by their plight. He will not, if his feelings are in a particular case left untouched, or more generally he finds his concern less and less liable to being caught, simply be unmoved by this. He will regret this, perhaps blame himself or feel remorse for allowing his heart to harden.\textsuperscript{272}

I have highlighted in the preceding passage what I take to be the as yet unsupported aspects of Dent’s account of the moral psychology of the virtues. He has argued correctly that the virtuous agent must do the things or have the properties he describes here. However, he has failed to give an account of the psychological structure of virtue. He must provide account of what he means by committed intent and how it is achieved, an account of how an agent resists alteration in his tendency to be moved by things of value, an account of what it is for agents such as human beings to be moved by value, an account of how agents like human beings go about ordering their thoughts, expectations, hopes, and fears to achieve stability of their dispositions, an account of what stability and constancy consists in, and an account of why regret and remorse are so significant for creatures such as us. Dent must give an account of the mental apparatus or “state” that results in virtue. He needs an adequate account of the stability and firmness of character and of how character (not direct and explicit deliberation) silences, screens, and alters our perception of salient features of our circumstances.

Someone who has the virtues, according to Dent, endeavors to suppress concerns and inclinations contrary to what is valuable, endeavors to stabilize and consolidate his or her liability to responsiveness to value, is liable to react with regret

\textsuperscript{272} Dent, \textit{The Moral Psychology of the Virtues}, 169 (emphasis added).
and self-blame if he or she is left unmoved on a particular occasion, (or if, more
pervasively, one’s heart becomes hardened), and is capable and ready to do what is
appropriate should any spontaneous inclination to do what is appropriate is absent.
These features represent ‘committed intent,’ which is a second-order desire273
concerning sustaining and regulating patterns of first-order desire. Dent argues that
the manifestation of ‘committed intent’ as a second-order desire will be in the form of
a set of ‘supports’ in the form of ‘potentialities’ to certain other actions and feelings
when first-order desires fail to be sufficient for right conduct. Since these
‘potentialities’ exist as part of an agent’s character, it is possible to say that when
these potentialities are actualized the agent acts from choice. We are not forced to
say, as it seemed we were earlier, “that whenever he acts he there and then engages in
prior deliberation and makes the choice of doing this rather than that. It could be in
the agent’s general unwillingness to allow his habit of desire to lapse, or in his
tendency to self-blame if he is left unmoved on a certain occasion, that the fact that
some desire of action of his constitutes his deliberate choice consists.”274 Therefore,
the natural conclusion to draw is that, in these examples, the potentialities that act as
supports of right action are the agent’s general unwillingness to allow his habit of
desire to lapse and his general tendency to self-blame. To illustrate how these
potentialities operate in virtuous activity, Dent examines bravery and says, “Bravery
is a virtue, and in virtue we choose how we shall act -- that is, decide upon acting
after previous deliberation. What Aristotle does not, perhaps, always make wholly

273 Dent borrows the notion of second order desires from Harry G. Frankfurt’s “Freedom of
the Will and the Concept of a Person” reprinted in The Importance of What We Care About (New

clear is how precedent deliberation may relate to a whole pattern in action, feeling, etc. over a long period.” Here Dent is emphasizing the temporally extended agency of human beings.

The problem is that Dent has not given an account of how his ‘potentialities’ operate in the moral psychology of the virtues. Nor does he give an account of how prior deliberation can affect current action. I think it is clear that he wants a way of describing how prior evaluative judgments act as inputs into practical reasoning in a way that shapes and focus deliberation at the moment of action. The planning theory provides an account of how this is possible and of how it occurs. Without an account of these potentialities or of how prior deliberation affects present deliberation, Dent cannot give a proper account of “committed intent.” Without an account of “committed intent” he cannot provide an account of how an agent resists alteration in his tendency to be moved by things of value, an account of what it is for agents such as human beings to be moved by value, an account of how agents like human beings go about ordering their thoughts, expectations, hopes, and fears to achieve stability of their dispositions, an account of what stability and constancy consists in, and an account of why regret and remorse are so significant for creatures such as us.

According to Dent, traits of character are the result of the connections and relationships between the “active powers” of an agent. An account of traits of character will be an account of these connections and relationships. The arguments presented above I think show that Dent has failed to give an account of these

---

connections and relationships. However, the temporally extended planning theory of character can fill these gaps.

5.6 Conclusion

In this chapter I addressed a potential objection to my strategy of deploying the temporally extended planning theory of human agency in the development of a moral psychology of character. The objection was that intentions (and policies) cannot be identified with commitments. However, we found no good reason for accepting the claim that the identification cannot be made.

I also argue that in addition to providing a coherent means of accounting for the functional roles of the virtues in practical reason and the moral life, the temporally extended planning theory of the moral psychology of character is superior to the moral psychology of virtue and vice provided by N. J. H. Dent in *The Moral Psychology of the Virtues*. The reason the planning theory account of moral psychology is superior is it can give an account of the concept of “committed intent.”
CHAPTER 6

MEETING ANSCOMBE’S DEMAND

6.1 Introduction

In this dissertation I was interested in understanding how it was possible for human beings to be virtuous. I have argued that we can understand virtuousness as the combination of a temporally extended planning theory of the moral psychology of character and a good-based account of objectively appropriate reasons for action. One might object that the account of being virtuous I have provided is too demanding, too elitist, and too intellectualist. Perhaps these objections have merit. However, I’m inclined to think they do not. They do not have merit because they imply that being virtuous is easy or at least that it is not exceptionally difficult.

Being virtuous is exceptionally difficult. A burgeoning literature stemming from the results of the experiments of social psychologists purports to show that there is no evidence that people possess distinctive character traits that could help explain differences in behavior. However, the findings of social psychologists do not tell us anything we did not already know. Human beings are weak and easily manipulated, particularly in stressful situations. These facts are so obvious as to be of little interest. If we are shocked by the numbers of individuals who act in morally

\textsuperscript{276} For a good discussion of the relevant literature and a rejection of the conclusion that being virtuous is impossible, see Christian Miller, “Social Psychology and Virtue Ethics,” \textit{The Journal of Ethics} 7 (2003): 365 – 392.
unacceptable ways, we are either naive or ignorant of the lessons of history. What is of interest is that some individuals hold up under stress and pressure.

In a recent book, Philip Gourevitch describes the Rwandan genocide which resulted in the deaths of an estimated 800,000 Tutsis between April and July, 1994.\textsuperscript{277} The genocide was driven by the Hutu Power movement which sought to eliminate from Rwanda the supposedly racially distinct Tutsi minority. For decades prior to the 1950s, the Hutu majority had been economically and socially subjugated by the Tutsi minority and their European backers. Despite a transition of power to the Hutu majority in the 1950s, racial animosity introduced by the promulgation of false European race science continued to plague Rwanda. Hutu dictators and their cronies encouraged the dissemination of propaganda concerning the political ambitions of the Tutsi minority, thereby sowing the seeds of evil among the Hutu population. After the assassination of the Rwandan President in 1994, racial hatred and political propaganda united to ignite one of the most efficient blood-letting of the Twentieth Century.

In the second quarter of his book, \textit{We Wish to Inform You that Tomorrow We Will Be Killed with Our Families}, Gourevitch provides a fascinating juxtaposition between a pair of active resistors to and a pair of complicit observers of the genocide. We can, I think, learn something of value concerning the moral psychology of character by examining how these individuals reacted under the conditions of the genocide.

\begin{footnotesize}
\begin{enumerate}
\item Philip Gourevitch, \textit{We Wish to Inform You That Tomorrow We Will Be Killed With Our Families}, (New York: Picador, 1998).
\end{enumerate}
\end{footnotesize}
Thomas Kamilindi was a radio journalist who refused to participate in the spread of Hutu Power propaganda and quit his job in protest. He put his life at risk by refusing to be complicit in the genocide. After the genocide began he was ordered to return to work at the radio station but refused and then was taken by soldiers to be executed only to be saved at the last possible moment by an army officer whom he did not know. Gourevitch “was struck that each of [Thomas’s] encounters with Hutu Power had followed a pattern: when the [Information] minister ordered him back to work, when the soldiers came for him, and when they told him to sit in the street [to be executed], Thomas always refused before complying.”278 After narrowly escaping death and agreeing to do one radio broadcast, Thomas know he could not be a party to the continuing genocide. Therefore, he sought refuge in the Hotel des Milles Collines.

Paul Rusesabagina was the manager of the Hotel des Milles Collines in Kigali, Rwanda. Rusesabagina used his hotel to shelter Tutsi refugees fleeing the killing squads of the Hutu Power movement. Gourevitch describes Paul this way:

Paul is a mild-mannered man, sturdily built and rather ordinary-looking – a bourgeois hotel manager after all – and that is how he seemed to regard himself as well, as an ordinary person who did nothing extraordinary in refusing to cave in to the insanity that swirled around him. “People became fools. I don’t know why,” he said to me. “I kept telling them, ‘I don’t agree with what you are doing,’ just as openly as I’m telling you now. I’m a man who’s used to saying no when I have to. That’s all I did – what I felt like doing. Because I

278 Gourevitch, We Wish to Inform You, 123.
never agree with killers. I didn’t agree with them. I refused, and I told them so.”

Self-reports are often unreliable. We have no way of knowing whether Paul is being modest or self-aggrandizing. We do know what he did. He continually refused to hand over the Tutsi refugees holed up in his hotel while using the resources of the hotel, particularly its liquor cabinets, to placate those who came to his door and demanded that his guests be handed over.

Father Wenceslas, on the other hand, was complicit in the slaughter of Tutsi refugees who sought shelter and protection in the Catholic cathedral, Sainte Famille, in Kigali. Later, in 1998, French authorities charged Father Wenceslas with “providing killers with lists of Tutsi refugees at his church, flushing refugees out of hiding to be killed, attending massacres without interfering, sabotaging…efforts to evacuate refugees from the church, and coercing refugee girls to have sex with him.” When asked to account for his actions during April, 1994, Wenceslas responded by saying, “I didn’t have a choice…. It was necessary to appear pro-militia. If I had had a different attitude, we would have all disappeared.”

Bishop Misago of Gikongoro gave answers similar to those of Father Wenceslas when he was asked about his silence in the face of genocide and his role in the killing of eighty-two school children. For example, he said, “I don’t have an

279 Gourevitch, *We Wish to Inform You*, 127.


army. What could I do by myself? Nothing. That is elementary logic.”

When asked why he did not try to use his influence to protect those he could, he replied by saying, “When men become like devils, and you don’t have an army, what can you do? All paths were dangerous. So how could I influence?”

What is so striking about these cases is the similarity of the responses given by each man concerning his actions during the genocide. Each of them spoke as if he had no choice. Thomas and Paul could not participate or be complicit in genocide. However, neither of them spoke of an obligation not to be involved. Instead, they spoke as if there were an inner compulsion not to be involved and to resist. For Paul Rusesabagina the issue of killing, or being complicit in killing, “posed no great challenge.” He was “surprised” that he was the exception, that he was a member of shockingly small minority of human beings who refused to take part or to be complicit in the slaughter of other innocent and helpless human beings. The possibility of participation was utterly unthinkable. For this reason, Paul refused to regard himself as “righteous, except when measured against the criminality of others, and he rejected that scale.”

While Paul might not agree that he was strong, preferring to regard his countrymen as weak, he was strong. And his countrymen were weak. “Many Rwandans didn’t agree with the genocide, of course, but many overcame their

282 Gourevitch, We Wish to Inform You, 138.
283 Gourevitch, We Wish to Inform You, 139.
284 Gourevitch, We Wish to Inform You, 141.
285 Gourevitch, We Wish to Inform You, 141.
disagreements and killed, while many more simply saved their own skins.” At the very least, Wenceslas and Misago were interested in saving their own skins. But interestingly, they too felt compelled to act as they did. Each of them spoke of feeling helpless and unable to resist. Of course, they could have resisted. They may very well have died as a result, but they could have resisted. Instead, they were able to consider participation and complicity as admissible options, as options worth considering and choosing.

Like Gourevitch, I am not interested in what makes someone like Wenceslas or Misago weak; I want to know what makes someone like Thomas or Paul strong. I think I have provided a plausible explanation of the moral psychology of an agent like Paul. Paul was committed to humanity and to reason. His commitment to these values made the thought of actively engaging in genocide unthinkable. On the basis of his valuing humanity and reason, Paul formed a self-governing policy to regard the objectively appropriate reason of preserving innocent life as a justifying end in motivationally efficacious deliberation. His various desires are then considered in light of his self-governing policy, such that, for example, the preservation of his own life is given relatively little weight in his deliberation. Presumably other desires are also given little or no weight in is deliberation, such as a desire to make a profit. Given his commitments and his self-governing policies, Paul constructed a series of policies, plans and intentions which allowed him to remain strong and resist temptation, social coercion, and manipulation by government

286 Gourevitch, We Wish to Inform You, 127.
287 Gourevitch, We Wish to Inform You, 127.
288 See Gourevitch, We Wish to Inform You, 141.
propagandists. So, when killing squads arrived at his door, he was prepared to resist and capable of doing so. Furthermore, he had plans for protecting his charges. His plans involved bribes, negotiations, flattery, and emotional appeals. He also had plans intended to diffuse conflict, divert attention, and deceive.

The majority of Rwandans were weak, not because they did not judge genocide to be evil, but because they were not psychologically capable of resisting social coercion and manipulation. Their deliberative structure was such that killing the innocent was an admissible option, something to be considered, something to be chosen given the available alternatives. Other Rwandans perhaps considered genocide an objectively appropriate reason for action. While these individuals were certainly mistaken, they may very well have valued genocide and constructed self-governing policies designed to promote this justifying end. Given the organization which produced the high efficiency of the slaughter, it is reasonable to assume that at least some individuals did so regard it. An individual with highly developed deliberative structure with evil at its core is certainly a dangerous person. However, far more dangerous is one such person with the ability to manipulate millions of individuals who lack character, who lack the ability to resist, to refuse, or to say no. A cause of genocide was that the population did not reject out of hand the mere suggestion of killing their neighbors, friends, and families. Somehow the filters of admissibility which are a feature of a properly functioning deliberative structure did not filter out killing the innocent.

Unfortunately for humanity, most of us do not have character. While genocide might seem an extreme case, the importance of the lesson to be learned is
not diminished. We are all in danger of mistakenly supposing that we are men and women of character who could resist social coercion, manipulation, temptation, fear, or self-interest. However, our best defense against weakness resulting from a lack of character is to attempt to approximate the moral psychology of virtue. One important aspect of the moral psychology of virtue is the possession of character. We must value what is good, we must form self-governing policies concerning what is good, we must form policies and plans appropriately related to what is good (and to what is evil), we must regard ourselves as temporally extended agents who may regret our present actions, and we must allow our past deliberations to influence sufficiently our present actions.


-------- “Modern Moral Philosophy.” In Crisp and Slote.


-------- “Planning and Temptation.” In *Faces of Intention*.


-------- “Identification, Decision, and Treating as a Reason.” In *Faces of Intention*.
“Practical Reasoning and Acceptance in a Context.” In *Faces of Intention*.


“Taking Plans Seriously.” *Social Theory and Practice* 9, nos. 2-3 (Summer-Fall 1983): 271-287.


“Moral Beliefs.” In *Cahn and Haber*.


-------- “Virtue Theory and Abortion.” In Crisp and Slote.

-------- “Applying Virtue Ethics.” In Hursthouse, Lawrence, and Quinn.


-------- “What Can Moral Philosophy Learn from the Study of the Brain?”


Manninen, Juha and Raimo Tuomela. Essays on Explanation and Understanding.


Novacek, Jill and Richard S. Lazarus. “The Structure of Personal Commitments.”


Searle, John. “How to Derive an Ought from an Is.” In Cahn and Haber.


-------- “Agent-Based Virtue Ethics.” In Crisp and Slote


-------- “Responsibility for Self,” In Watson, Free Will.


Trianosky, Gregory Velazco y. “What is Virtue Ethics All About?” In Statman.


Watson, Gary. “On the Primacy of Character.” In Flanagan and Rorty.


-------- “What is Knowledge? (Draft)” published as “What is Knowledge?” In Greco and Sosa.