SELF AND WILL:
AN EXISTENTIAL THEORY OF MOTIVATION AND
FRANKFURT'S THEME OF IDENTIFICATION

A Dissertation

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

Doctor of Philosophy

by

John Joseph Davenport, B.A., M.A.

__________________________
Karl Ameriks, Director

Department of Philosophy
Notre Dame, Indiana

July 1998
SELF AND WILL:
AN EXISTENTIAL THEORY OF MOTIVATION AND
FRANKFURT'S THEME OF IDENTIFICATION

Abstract
by
John Joseph Davenport

This dissertation develops a new account of the will as a motivational power of commitment or resolve, which contrasts with volition in the `thin' sense as a mere decision to form an intention. After tracing the historical background of these concepts in Chapter I, Chapter II argues that will involves the capacity to motivate oneself "projectively" in devotion to causes and purposes for which one did not necessarily have an antecedent "desire." "Projective motivation" differs from three types of "desire" —(1) inarticulate appetites and generic urges, (2) subjective inclinations and preferences for definite objects, and (3) evaluative desires following from judgments that something is objectively `good' for us—all of which arise from the sense that something is lacking in the agent. Projections do not share this teleological lack-structure, and are grounded in considerations other than the agent’s well-being, such as the existential meaningfulness of pursuing ends, as distinct from realizing them.

This concept helps resolve problems with eudaimonist moral psychologies. It also helps make sense of Kant’s notions of spontaneity and autonomy, although moral motivation as Kant conceives it is only one form of projective motivation. Hence the usual teleological and deontic models do not exhaust all the kinds of motivation there are. Chapter III then argues that thinkers as divergent as Scotus, Levinas, Arendt, and Frankfurt implicitly appeal
to the possibility of projective motivation.

The last two chapters focus on Harry Frankfurt’s hierarchical account of volitional "identification" (a reflexive intrapersonal attitude towards certain emotions or desires taken as authoritative expressions of one’s self), comparing this notion to Heidegger’s and Foucault’s interpretations of ‘care of the self.’ Chapter IV distinguishes and reformulate five objections against Frankfurt’s initial account of identification in terms of second-order desires; these objections point towards the need for a qualitative distinction between ordinary desires and the states which constitute identification. It then analyzes Frankfurt’s later attempts to address these difficulties, and argues for a new account of identification in terms of ‘higher-order projections.’ Chapter V defends this new existential analysis against rival accounts of identification by Stump, Taylor, Watson, Velleman, Bransen, Buss, and Dennett.
For Robin,

who has heard it all
at one time or another.
# Table of Contents

This listing does not include subsections within the main sections of each chapter. Detailed contents for chapters are given at the beginning of each chapter.

[These are the page numbers in the official version; they are not all preserved in this version]

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface: The Project of an Existential Interpretation of Personhood</td>
<td>vii</td>
</tr>
<tr>
<td>Introduction: Existentialism and the Will</td>
<td>1</td>
</tr>
<tr>
<td>Int.1. The Classical Substance-Model of Personhood</td>
<td>2</td>
</tr>
<tr>
<td>Int.2. The Alternative Existentialist Conception of Persons: Heidegger</td>
<td>6</td>
</tr>
<tr>
<td>Int.3. Three Conditions for an Existentialist Theory of the Will</td>
<td>18</td>
</tr>
<tr>
<td>Chapter I: Historical Background: Two Concepts of the Will</td>
<td>36</td>
</tr>
<tr>
<td>I.1. The Role of Volition</td>
<td>38</td>
</tr>
<tr>
<td>I.2. The Heritage of the Thin Concept of Will</td>
<td>57</td>
</tr>
<tr>
<td>I.3. The Thick Concept of Will and `Middle Soul’ Motivation</td>
<td>72</td>
</tr>
<tr>
<td>I.4. Sentiment, Will, and the Ambiguous `Middle Part of the Soul’</td>
<td>78</td>
</tr>
<tr>
<td>I.5. Towards a Preliminary Moral-Psychological Taxonomy</td>
<td>88</td>
</tr>
<tr>
<td>Chapter II: Desire and Will:</td>
<td>92</td>
</tr>
<tr>
<td>The Structure and Existential Role of Projective Motivation</td>
<td>94</td>
</tr>
<tr>
<td>II.1. Introduction</td>
<td>95</td>
</tr>
<tr>
<td>II.2. Three Kinds of Desire</td>
<td>106</td>
</tr>
<tr>
<td>II.3. A Preliminary Interpretation of Desire as `Lack’</td>
<td>126</td>
</tr>
<tr>
<td>II.4. Nine Types of Second-Order Desire</td>
<td>138</td>
</tr>
<tr>
<td>II.5. Will as Conative Projection</td>
<td>173</td>
</tr>
<tr>
<td>II.6. The Existential Significance of Projection</td>
<td>183</td>
</tr>
<tr>
<td>II.7. Kant and Projective Motivation</td>
<td>201</td>
</tr>
<tr>
<td>II.8. Projection and Libertarian Freedom: A Concluding Note</td>
<td></td>
</tr>
<tr>
<td>Chapter III: Illustrations of Volition as Projective Motivation:</td>
<td></td>
</tr>
<tr>
<td>Frankfurt, Scotus, Levinas, and Arendt</td>
<td>215</td>
</tr>
<tr>
<td>III.1. Introduction: Five Key Points on Projection</td>
<td>217</td>
</tr>
<tr>
<td>III.2. Frankfurtian Care as Projective Motivation</td>
<td>218</td>
</tr>
<tr>
<td>III.3. Williams’s Neo-Humean Internalism</td>
<td>228</td>
</tr>
<tr>
<td>III.4. Duns Scotus on Free Will</td>
<td>231</td>
</tr>
<tr>
<td>III.5. Levinas on Superabundant Will and Volitional Generosity</td>
<td>239</td>
</tr>
<tr>
<td>III.6. Arendt’s Distinction Between Circular and Linear Motivation</td>
<td>248</td>
</tr>
</tbody>
</table>
Chapter IV: A New Phenomenology of Frankfurtian Volitional Identification .... 257
   IV.1. Introduction: The Importance of Identification .................. 259
   IV.2. Frankfurt’s Initial Account of Volitional Identification in
         Terms of Desires .................................................. 271
   IV.3. Objections to Frankfurt’s Initial Account ......................... 284
   IV.4. Developments in Frankfurt’s Later Accounts of Identification ... 310
   IV.5. A First Approach to Decisive or ‘Wholehearted’ Identification ... 337
   IV.6. The Projective Account of Identification ........................ 361

Chapter V: A Critique of Several Alternative Accounts of Identification .... 389
   V.1. Introduction: Categorizing the Alternatives ...................... 391
   V.2. Identification as Rational Evaluative Judgment .................. 393
   V.3. Decisive Identification as Self-Discovery: Bransen and Tugendhat 427
   V.4. Buss’s Nonattitudinal Account of Identification ................. 433
   V.5. A Critique of Dennett’s Neo-Behaviorist Account of Identification 443
   V.6. Conclusion: Volitional Identification and Personhood ............ 460

[Chapter Six on Freedom is included in this unofficial version]

Bibliography ................................................................. 472

Index ................................................................. 493
ACKNOWLEDGMENTS

Too many people to name have aided in the development of the ideas gathered together in this dissertation. Many of the points about identification and personhood were first sketched out during a graduate seminar taught by Dieter Sturma in the fall of 1993 during his semester at Notre Dame, and subsequently in conversations with Mitch Clearfield. The debt I owe to Harry Frankfurt’s writings is obviously also enormous, and, despite my disagreement with him on some of the most important questions, he has been a very helpful correspondent. Similarly, John Fischer and Eleonore Stump have also been patient in talking with me about several issues raised in the dissertation.

The existentialist ideas underlying the dissertation developed during several seminars and directed readings with Steve Watson, whose writings have also altered many of my opinions. Some of their original inspiration goes back to a class at Yale University led by George Schrader, one of those professors one never forgets. More recently, I have profited from a seminar on current issues surrounding free will led by Fritz Warfield. However, although its terminology reveals a Heideggerian pedigree, I feel that the intuition expressed in the theory of "projective motivation" has been with me almost as long as I can remember having any ideas about moral psychology. In fact, it is what I just took motivation obviously to be, until, upon starting to study philosophy, I realized that my notion was not at all standard. My sense of the need to hang onto this intuition, perhaps out of a spirit of resistance, was whetted by a number of classes covering topics in ancient and medieval philosophy, including one wonderful seminar on the will led by Mark Jordan. But although I take credit (and blame) for projective motivation, this idea is also undoubtedly due in part to the large influence that Stephen Donaldson’s first novels had on me when I read them in the
early 1980s.

My director Karl Ameriks has been more tolerant and helpful than I can say. He has helped me refine what was initially an ambiguous and overly ambitious plan to analyze personhood in general into a more coherent and focused whole. His particularly challenging questions on the issues related to Kant and free will are reflected several places in the dissertation. Along the way, I also appreciated his good advice on the more pragmatic and personal aspects of getting a job like this done in the right way. And most of all, I must thank him for his infinite patience with the changes this project has been through.

I’m also indebted to David Solomon both for his insights in several seminars at Notre Dame, and for his searching questions about my arguments. And I thank Tom Flint for filling in as a third reader for the dissertation at the last minute. I probably would not have graduated on time without his help, for which I will always be most grateful. Finally, thanks are owed to my parents and my wife Robin for tolerating and encouraging my passion for philosophical insight and my search for new perspectives.
PREFACE:

The Project of an Existential Interpretation of Personhood

The Issue. Although it remains popular among a wider public, enthusiasm for the existentialist approach to personhood has been declining in serious philosophical literature since the 1970s. In analytic philosophy, metaphysical writings on personal identity over time have dismissed existentialists contributions on the complex temporality of selfhood as obfuscation. Likewise, mainstream metaphysical authors have new ways of conceiving essence, and have difficulty in making sense of the existentialist claim that for persons, ‘existence precedes essence,’ unless this is read just as a rather confusing way of saying that we enjoy some sort of libertarian freedom (instead of as a metaphysical consequences from such freedom). Since the development of modal logic, debates about the metaphysics of free will have been rewritten in a language relative to which existentialist writings on freedom may seem outdated. Debates on whether moral responsibility requires any sort of libertarian freedom, as existentialists commonly held, have also become much more complex since Frankfurt’s 1969 presentation of putative counterexamples to the Principle of Alternate Possibilities. Among Neo-Kantians, compatibilist theories of autonomy have gained popularity, and their Neo-Aristotelian critics often regard existentialism as a final defense of enlightenment individuality that reduces the person to an isolated, ghostly, and arbitrary free will, an “unencumbered self” in Michael Sandel’s memorable phrase.

Developments in feminism and continental recent philosophy have re-enforced this criticism, arguing that persons are essentially social beings who can only understand themselves or even develop a ‘self’ in terms of their relations to others and the implicit
norms defining the sphere of activities in which they conduct their lives. In pragmatism and some forms of radical hermeneutics, the notion of personhood itself is treated simply as a social convention or an artifact we require as an underpinning for our moral and legal language games.

At the same time, existentialists are dismissed by psychologists for having too anti-naturalistic a notion of human motivational powers. In philosophy of mind, the debate is focused on whether the intentionality of mental states is something more than syntactic operations, and whether the sentience that computer-states would have to enjoy for computers to counts as persons is irreducible to physical properties of brain states. Will and freedom are not even at issue in these debates. In moral psychology, the dominant debate is between a range of neo-Humean and neo-Kantian positions, in which (to give a crude summary) all motivation terminates in desires we simply acquire naturally or accidentally, or some motivation ultimately stems from a choice to follow neutral rational judgments. These extremes leave no room for the rich picture of self-motivational capacities that existentialist writers have painted.

My dissertation is the beginning of an attempt to revive existentialism by addressing these issues. The idea is eventually to present a conception of personhood that is recognizably existentialist, or similar to that presented by writers like Kierkegaard, Heidegger, Jaspers, and Sartre in certain fundamental ways, but that takes into account the last twenty years of developments in the many different areas of philosophy that directly affect our understanding of what it is to be a person. The result will hopefully be a more ‘sophisticated’ existentialist theory of personhood that can be presented in contemporary terms as a serious challenge to current dogmas in metaphysics and moral theory, and be
defended against the ascendant naturalistic, Humean, rationalistic, compatibilist, or pragmatist alternatives.

The dissertation can only complete part of this agenda, and it focuses on the crucial first step of developing an adequate conception of willing. In accord with the goal of making existentialism contemporary, my strategy is to take advantage of Harry Frankfurt’s insight that persons are distinguished from other animals by their unique motivational abilities (an insight developed in work that is not about the libertarian/compatibilist debate). Chapters I and II develop the idea that the will characteristic of persons is not simply the formation of an intention or a choice between different possible actions, but a commitment or resolve of the whole person in which we steel ourselves to some task, purpose, plan, or goal, despite its difficulties or risks. As I interpret it, such resolve is itself a kind of motivation unlike other sorts, especially those typically called "desires." There are several different sorts of desire, but they share in common the feature that the person is moved by the attractiveness of an object to desire it, whereas in what I call projective motivation, we are able to give ourselves ends which we did not necessarily desire to pursue. The evidence I cite for projection in this volitional sense suggests that some motivation itself is something in which the agent participates actively, rather than being merely its passive recipient. But this is not meant to imply that persons just arbitrarily or for no reason hurl themselves towards some end; there are always grounds for projections, i.e. there is always a story to be told about why an agent projected some end for herself.

The second half of the dissertation also builds on Frankfurt’s work, in particular his claim that persons are distinguished by their capacity to be concerned about and "identify with" their own motives for acting. Chapters Four and Five focus on this reflexive case of
adopting ends concerning our own character, showing that it sheds light on existentialist pictures of selfhood, and that what is happening when we take a stand for or against particular motivating dispositions, desires, and emotions can best be explained in terms of the notion of projective motivation developed in the first half. There are two interlocking claims here: (1) the hypothesis of projective motivation allows us to resolve longstanding problems concerning Frankfurt’s own initial explanation of reflexive identification with motives, and (2) the manifest phenomena of intrasubjective volitional identification are thus more evidence for the existence of projective motivation.

This analysis of identification also provides a key component for the new existential conception of personhood: it allows us to see how persons are involved in shaping their own character or ‘self.’ On this picture, the freedom to form ourselves is not an abstract speculative detachment; rather, it is a freedom of our reflexive motivational capacities, i.e. a liberty to "identify" with some of our possible motives for action and to reject others. Through such volitional identification, the person motivates herself to care about the shape and content of her own motivational character.

This analysis ultimately points towards a virtue interpretation of personhood, according to which a person is not just a being capable of moral responsibility for her actions, but also a being capable of shaping her own moral character corruptly or virtuously—an entity capable of being inwardly good or evil. This virtue interpretation picks out the intuitive subject matter of the concept to be explained in any theoretical conception of moral personhood: what is distinctive of persons.

However, it is possible to pursue either of aspect of the virtue interpretation of personhood—character and freedom—while ignoring the other. For example, in After
Virtue. Alasdair MacIntyre equates the personal individual with the overarching "character" that makes the narrative of his actions and practices intelligible to others in the scope of his or her complete life. As in a drama, in which each actor is cast as a certain character — a character whose characteristics they can interpret and vary to some degree but not completely rewrite — at the deepest level, each of us is a certain character that we cannot change, since it is the root of all of our active engagements and desires for self-development.

By contrast, Jean-Paul Sartre analyzes the ‘person’ as the for-itself of consciousness, which ultimately ‘is’ only its freedom, or transcendence of its factical being. Because of the absoluteness of this freedom, the for-itself cannot be any ‘character’ the person acts out in social interactions: such characters can only be roles in which we lose ourselves in bad faith. As a result, the attempt to structure our identity in terms of values (such as concepts of virtue) is a distortive attempt to mask the nothingness that separates us from our social roles, and is bound to fail in "anguish." On this theory, the person cannot be identified with character-defining values in anything but a self-deceptive sense.

The Structure of the Argument. Unlike either MacIntyre or Sartre, an existentialist virtue theory of the person must avoid both the fallacies of narrative identity without freedom and freedom without substantive character by showing how volitional character and the freedom of the ‘I’ are interconnected in the existential self. This project requires three main steps. The first step, which is the task of this dissertation, is to develop a phenomenology of the will, or to analyze the volitional dimension of personhood. This analysis will support the existentialist contention that intrapersonal or subjective dimension
of personhood is not primarily constituted simply by self-consciousness or our cognitive capacities for practical reason, but rather by the ability to care for the self, as Foucault and Heidegger describe. To explain this capacity, the first two chapters of the dissertation will provide a new analysis of volition as a substantive motivational concept; the third chapter will discuss several philosophical treatments of volition that illustrate this idea of ‘projective’ self-motivation.

The fourth chapter will then turn to Harry Frankfurt’s argument that persons are distinguished from other animals by their power to form "second-order volitions" through which they identify with different possible motivations or reasons for acting. This chapter gives a phenomenological reconstruction of Frankfurt’s insights that puts them on firmer footing, drawing on the prior analysis of will as projective motivation. This reconstruction will (a) show that ‘intrasubjective volitional identification’ is distinct both from appetite or subjective preference and from ‘neutral’ ethical judgments about what is valuable or best; (b) develop Frankfurt’s notion of ‘care’ into a theory of volitional character.

The fifth chapter will defend this analysis of volitional identification against several rival interpretations. I critique rationalist accounts of identification offered by Eleonore Stump, Charles Taylor, Gary Watson, and David Velleman. I then turn to Jan Bransen’s hermeneutic interpretation of decisive identification, which is based on work by Ernst Tüngendhat, and Sarah Buss’s critique of the idea that there is any attitude or state constituting identification in Frankfurt’s sense. This chapter concludes with a critique of Daniel Dennett’s pragmatist reinterpretation of volitional identification, and in this light, argues that when they are properly understood, the implications of Frankfurt’s work for our conception of personhood are more radical than has been recognized.
The dissertation was supposed to conclude with a final chapter arguing (a) that the capacity for volitional identification necessarily includes libertarian freedom, or the practical ability to shape one’s inner character in alternate possible ways; (b) that there are no true Frankfurt-counterexamples to a principle requiring liberty of identification for responsibility for character; and (c) that an existentialist model of factically constrained libertarian freedom can be defended against a series of familiar objections. Although complete, this work had to be left out of the dissertation for reasons of space.

The Larger Project of which the dissertation is part. This is as far as the dissertation goes in developing an existential conception of personhood. To complete this account, four other steps are left for future research. I will review them briefly in order to give the reader a better sense of the larger picture into which the dissertation’s analysis of volitional motivation and self-formation will eventually fit.

The first of these three uncompleted steps is to trace the historical development of the idea of freedom from Kant, who already gives it a minimal practical form distinct from mere spontaneity, through Schelling to Kierkegaard and Heidegger, who give liberty a progressively more substantive embodiment in our motivational character and temporal experience. This ‘schematization’ of raw liberty will verify the existentialist credentials of the account of free will sketched in the last chapter of the dissertation, and help fill out that account of volitional liberty as non-teleological yet immune to the problem of arbitrariness objections. A closer look at examples from existentialist literature will clarify the sense in which personal liberty is not a bare unstructured spontaneity, but rather a freedom of the temporal will—a freedom that has both a dispositional form and a history for each
individual.

The second remaining step is to develop an existential account of the relations of persons to both the world, or Being in general. This ontological analysis involves a review of the alternative paradigm ‘between’ scientific realism and Berkelian idealism inherited by 20th century existentialists from the phenomenological school,¹ and a revision of Nagel’s account of how the "objective" and "subjective" perspectives fit together.

Third, I need to explain how the reconstructed existentialist analysis of personhood fits with the theory of consciousness emerging from ‘Heidelberg School’ authors such as Dieter Henrich and Manfred Frank, who argue for the irreducibility and pre-reflective character of first-personal self-awareness. Although this account fits with the existential account of the relation between person and world, we ought not to explain personhood simply by reference to the existence of this form of pre-thetic self-awareness, because this is the condition of possibility for consciousness in general.² Instead, personal self-consciousness is distinguished by a stronger sense of ownership of one’s own mind, which is primarily a result of the capacity for volitional identification. The structure of the will discovered in this dissertation will show that a person is always comported practically towards the actual contents of her consciousness and their other possible contents. In part, this is because our factual and counterfactual states of motivation, such as impulses, preferences, desires, and associated maxims for action are encountered as possible subjects

¹I have begun this work in my paper, “Heidegger’s Phenomenological Argument Against Representationalism,” presented at the Pacific Division Meeting of the American Philosophical Association (Los Angeles, CA, March 28, 1998).

²I have treated this topic in my paper, "Churchland’s ‘Rediscovery’ of Analogy-Argument Fallacies: The Irreducibility of Prerreflective Self-Awareness,” presented at the Pacific Division meeting of the American Philosophical Association (Berkeley, CA, March, 1997).
for volitional identification. In personal self-awareness, consciousness is appropriated to a self because it serves as a kind of material base for the activity of intrasubjective volitional identification, which together provide the basic structure of the ‘personal world.’

Finally, as mentioned at the start, my existentialist account of personhood must be defended against the dominant rival paradigm today, which embraces a broad family of interpersonal conceptions of personhood, such as those presented by Hegel, Max Scheler, Edmund Husserl in *Ideas II*, Jürgen Habermas, and Alasdair MacIntyre. What unites their accounts is the thesis that personal identity requires a kind of practical unification of the individual over time through character, which itself depends on the rational capacities of persons for shared horizons of meaning, relative to which any volitional self-relation must also be derivative. Contrary to existentialism, however, all these accounts imply that there is one fundamental and unique ‘character’ which is the ‘self’ that individuates a single person, since it accounts for the unity of his moral selfhood over time; since anything belongs to the person by virtue of its relation to this unique narrative "style," this character-‘self’ must by definition be unchanging. I regard this kind of ‘narrative essentialist’ conception of the person (which Paul Ricouer also shared at one time) as a modern hermeneutic recasting of the Boethian concept of the person as individual rational substance. Just as the Boethian conception gives us a transcendent motivational telos that we cannot reject or choose, narrative essentialism maintains that our deepest level of ‘narrative unity’ is ultimately outside our control.

Against this view, an existentialist conception maintains that we cannot be persons unless we to some extent transcend of our character, and thus have the capacity to change ‘who we are’ even at the deepest level of selfhood. Like the Molinist possible person, whose
The stress on the word 'individual' here is meant to indicate that this statement, which I think is the best way of construing the basic thesis of existentialism, does not commit us to saying that there are no structures of personhood in general. What the existentialist by definition rejects is the idea that the morally judgeable character of any person amounts to an individual substance in the sense of a maximal set of features that make her what-she-is, i.e. a universal with one instance whose instantiation is the actualization of this possible person. An existentialist maintains, on the contrary, that the person whose part we play cannot ultimately be fixed in this way, since our liberty reaches to the very roots of our character, though it also always exists in time and already has some accomplished shape and content. The person is not defined by an individual essence, narrative or otherwise, that logically or temporally precedes her own existence.

This is not to say, however, that interpersonal relations are not essential to personhood. The final account will have to explain how our reasoning capacities, which are inherently social, are equiprimordial with (and linked to) the volitional capacities on which this dissertation focuses. Once all these tasks are carried through, we will have a complete existentialist account of all central aspects of personhood, which has been defended against the most important rival views. Although the dissertation only takes us the first part of the way towards this goal, at various points there are anticipations of what this completed existentialist conception of personhood will look like.

Methods. I should also say something about my philosophical methodology in addressing these issues. Although I employ some of the same analytical and historical tools
as others writing on my topics, my method is also broadly speaking \textit{phenomenological}. Although this is not generally in the foreground of my discussion, it becomes important at some crucial junctures in the argument, including Chapter I §5, Chapter IV §3.1, and Chapter V §5.

In general, by a phenomenological approach, I mean one that distinguishes between the primary phenomena to be explained in some area of philosophy, and the rival theoretical explanations that construe these phenomena in different ways. The phenomenological approach presumes that we usually can discern, however imperfectly, some important phenomena that serve as paradigm cases or fixed points of reference for a particular problem; the task of theoretical explanation is then to provide as convincing an account of these phenomena and their grounds of possibility as can be given, where what counts as "convincing" is \textit{itself guided by the shape of the phenomena that present themselves} more or less clearly in common human experience. Thus the first aim of theoretical explanation is to follow where the phenomena lead rather than to make them fit the mold of a metaphysics to which one is antecedently committed. This principle, which corresponds in form to the Habermasian communicative ideal of reaching conclusions based solely on \textit{the force of the better argument} alone, is important in my case for deciphering how we can even begin to analyze concepts such as the will, volitional identification, and freedom.

But of course I am aware that objections have been raised against this principle. Let me mention and briefly respond to three. (1) \textit{The hermeneutic objection}: the ‘phenomena themselves’ are never pure givens; our reception of them is mediated by a host of unexamined presuppositions (some of them theoretical and even metaphysical), which vary both culturally and historically, and it could not be otherwise for beings like us. (2) \textit{The
Linguistic objection: our evaluation of whether an explanation meets, satisfies or (in older lingo) adequately saves the phenomena is always mediated by linguistic structures whose implications exceed our capacity to make them certain beforehand in reflection, and which also vary over time. (3) The Underdetermination objection: two theories may save all the phenomena equally well, leaving us to decide between them on other criteria.

In my view, the caveats expressed by (1) and (2) only show that judging an explanation’s convincingness according to the pure phenomenological approach (or philosophical ‘strict science’ in Husserl’s sense) is always a counterfactual ideal, not that we should not try to approximate this ideal as best we can, nor that we have no ways to tell when we have done a better or worse job at approximating it. We cannot spell out a method that is agreed on in advance to resolve disputes about the content, relevance, and reliability of our phenomena between parties in different traditions and cultures, but the process of spelling out rival descriptions usually reveals differences in quality of interpretation that would be apparent to neutral observers, were there any. And the problem that we cannot ever be entirely neutral observers is itself revealing for several issues in philosophical anthropology and epistemology.

Objection (3) poses different problems, but for the most part, in this problem only becomes serious in the philosophy of science and the philosophy of physics in particular; in moral philosophy and philosophical anthropology, we never get theories that clearly save all the most relevant and reliable phenomena, and so never get ones that do equally well. The problem is more one of judging between different imperfections.

Notations. Footnotes are numbered sequentially within each chapter to avoid very
high number notes. I use double quote-marks to indicate actually quoted text or the phrases used by other authors. I use single quote-marks when I am *mentioning* a term, or as scare-quotes for an expression or an assumption that I do not necessarily agree with. In places, I have also used subscripts to facilitate keeping track of the different *orders* of motivation being discussed.

I have tried to give a full reference for each article and book in the footnotes the first time they are cited, in addition to a full reference in the bibliography. I have left out the publisher’s town or city location in most references, since now that most publishers publish in multiple locations, in my view the practice of referring to the location of publication is outdated.
CHAPTER I

Historical Background: Two Concepts of the Will

How far from both muscular heroism and from the soulfully tragic spirit of unselfishness which unctuously adds its little offering to the sponge cake at a *Kaffe-Klatch*, is the plain simple fact that a man has given himself completely to something he finds worth living for.\(^4\)

Contents of Chapter I

I.1  The Role of Volition

I.1.1. Thick and Thin Concepts of the Will
I.1.2. ‘Volitionalism’
I.1.3. Voluntariness and Autonomy

I.2.  The Heritage of the Thin Concept of Will

I.2.2. Hobbes, Hume, and Others
I.2.3. James
I.2.4. Wittgenstein and O’Shaughnessy
I.2.5. Other Recent Treatments of the Will on the Thin-Concept Paradigm

I.3.  The Thick Concept of Will as ‘Middle Soul’ Motivation

I.4.  Sentiment, Will, and the Ambiguous ‘Middle Part of the Soul’

I.5.  Towards a Preliminary Moral-Psychological Taxonomy
I.1. The Role of Volition

I.1.1. Thick and Thin Concepts of the Will

In different areas of philosophy and psychology, from analytic philosophy of mind to continental ethics, from personality psychology to clinical psychotherapy, the idea that "the will" as a special capacity distinguishing human persons from other animals is out of fashion among most writers today. Although they may find harmony in little else, a majority of theorists in these disparate domains agree on refusing a role or place for volitional phenomena in any sense irreducible to other ‘folk’ concepts, such as belief, desire, disposition, behavioral pattern, interpersonal style, rationalization, mood, and so on. Across the spectrum, from theories of motivation and moral responsibility to the metaphysics of freedom, the concept of ‘will’ is much maligned as an outdated artifact, a scholastic ‘faculty’ that has become obsolete in our more subtle accounts of intentional action, rationality, autonomy, valuation, and character. For example, in psychology, personality types are defined in terms of behavioral traits, interests, interaction styles, and emotional dispositions, without any reference to volition.\(^5\) Meanwhile, in continental philosophy since Heidegger, for the most part ‘will’ has either been celebrated as an unconscious drive, a blind Nietzschean "will to power," or (partly because of these Nietzschean overtones) it has been rejected as unethical ‘willfulness,’ i.e. possessive, totalizing, domineering self-assertion, reckless of our relations to others and inimical to caring attentiveness.\(^6\)

---

\(^5\)See, for example, the theory outlined by such leading personality psychologists as Hans J. Eysenck and Michael W. Eysenck in their study, *Personality and Individual Differences: A Natural Science Approach* (Plenum Press, 1985).

\(^6\)Yet, as we will see in Chapter III, Levinas is a spectacular exception to this trend which Heidegger perhaps more than anyone else helped to foster.
For example, in a creative recent book, Jean-Luc Nancy argues that the freedom which is "the essential fact of existence" cannot be understood as the correlate of any kind of necessity known to philosophy, or explained in terms of any modal concept of possibility or causality, for this always turns freedom into the "free will" that makes the subject master of its representations. Like the existentialists preceding him, Nancy is opposed to the idea that freedom is just a property that persons possess, or something that entities simply exhibit or instantiate if they are individual subjects. So he follows Heidegger in rejecting what he understands as the "voluntaristic will" and all theoretical explanations of freedom, which only block our experience of freedom as a practical openness or "free space" for existence.

There are very different reasons for the decline of the will in other disciplines. In psychology, Freud’s influence is undoubtedly the greatest factor. As the prominent existential psychotherapist Irwin Yalom says,

Freud’s model of the mind...was based on Helmholtzian principles—that is, was an antivitalistic, deterministic model where the human being is activated and controlled by "chemical-physical forces reducible to the force of attraction and repulsion." Freud was unrelenting on this issue....Behavior is a vector, a resultant of the interplay of internal forces.

In other words, Freud’s conception of motivation was fundamentally Hobbesian, the result

---


8Ibid, p.36.

9Ibid, p.5.


11Ibid, p.18. By this, Nancy means a will that simply *represents* various possibilities to itself and then intends one of them. Instead, existence is a state lacking "an essence" that could be "represented or intended," and so its possible ways of being are not given to it beforehand: they do not even become *intelligible* to the person except through its free decision to posit them. So will is the way that existent beings go out of themselves towards what is unknown or unanticipatable: existential will makes its end or goal meaningful only through commitment to it. As we will see, however, this view is simply a more radicalized version of the conception of will as projective motivation.

12Ibid, pp.40, 43.

of a contest between inclinations which themselves ultimately trace to pain and pleasure stimuli. In the *Leviathan*, after having defined "deliberation" as the "alternate Sucession of Appetites, Aversions, Hopes, and Fears...no lesse in other living Creatures then in Man,"\(^\text{14}\) Hobbes says:

In *Deliberation*, the last Appetite, or Aversion, immediately adhaering to the action, or to the omission thereof, is that we call the Will; the Act, (not the faculty) of Willing. And Beasts that have Deliberation, must necessarily also have Will. The Definition of Will, given commonly by the Schooles, that it is a Rationall Appetite, is not good. For if it were, then could there be no Voluntary Act against Reason. For a Voluntary Act is that, which proceedeth from the will, and no other....Will therefore is the last Appetite in Deliberating. And though we say in common Discourse, a man had a Will once to do a thing, that nevertheless he forbore to do; yet that is properly but an Inclination, or Appetite.\(^\text{15}\)

The will is not a different kind of agency standing above our appetites and aversions; it is simply whatever desire emerges as strongest in the ‘economy’ of our attractions and repulsions. As a result, there is no sense in which we can act voluntarily out of appetite but against our real will, since there is nothing more to our will than the appetite on which we act. Although Hobbes’s account is reductive in the further sense that it holds that all intentions arise must from desires rooted in sense experience, it is similar to opposed ‘externalist’ theories (which allow intentions to flow directly from nonemotive cognitive evaluations or impartial practical reasons) in equating the will with the final operating intention.

It is this reduction of will to the intention which guides action (especially bodily performances) that has led to the declining significance of volition in twentieth-century philosophy of mind and moral psychology. Hobbes and his empiricist followers bequeathed

---


\(^\text{15}\)Ibid, p.127-8.
to contemporary philosophy the prevailing assumption that ‘volition’ is equivalent to whatever conditions make human actions voluntary or imputable. The will is thus reduced to what Pink calls "first-order agency," or the capacity to perform actions —doings with a purpose explained by attitudes "such as desires and intentions"— that are not themselves aimed at forming or sustaining such action-explanatory attitudes.\(^\text{16}\)

Yet in nonphilosophical discourse or ordinary language, "will" retains its richer connotations. As Pink has argued, in ‘ordinary psychology,’ the decision-making by which we form intentions to act and apply reasons to plan future actions itself counts as a special type of intentional action. Thus volition has the ‘thick’ or substantive sense of second-order agency. But "will" signifies more than a reason-considering decision to act in some definite way or other: in addition, it also connotes a distinctive kind of motivation for such ‘second-order’ acts through which we form first-order intentions. Specifically, "will" in its most familiar usages is associated with "determination," "diligent purposiveness,"\(^\text{17}\) and making difficult decisions; it signifies a way of focusing one’s energies of focus on doing something, or inward resolve to pursue a certain course, which is especially apparent in the face of adversity. If will is understood in this ‘thick’ and motivational sense, then it follows that many of our voluntary acts may occur without any motivation of the distinctively volitional kind at all: voluntariness by itself is a weaker condition than volitional commitment. Of course this does not mean that distinguishing voluntary actions from involuntary performances is not a weighty philosophical problem, or that we shouldn’t be held legally liable for everyday acts (such as brushing our teeth, or unselfconsciously humming while we work) that are often unrelated to any volitional resolve. Rather, the distinction between

\(^{16}\)Pink, \textit{The Psychology of Freedom}, p.17.

voluntariness and volition in the ‘thick’ sense opens up the possibility that many voluntary acts do not profoundly affect the shape of our character or its moral worth in the same way as volitions do.

However, this familiar thick motivational sense of volition should not be restricted, as Iris Murdoch suggests, "to cases where there is an immediate straining, for instance occasioned by a perceived duty or principle, against a large part of preformed consciousness." The motivational quality named by "will" does not always require a sense of inner tension, or an inward ‘gritting of mental teeth,’ as when Whoopie Goldberg struggles to hand over her check for $4 million to the Salvation Army nuns in the film *Ghost*. Volition is just as much in evidence when someone commits himself, without reservation, to doing something he already had other reasons for wanting to do, such as joining a team whose other members already like him.

Neither does the thick motivational sense imply that volition is ever unrelated to action. On the contrary, any psychic state (perhaps such as a ‘mere wish’) that does not look towards our real possibilities of acting, either because it moves us to act now or would so move us if the conditions made the relevant acts causally possible, does not meet a necessary condition implicit in the ordinary or everyday sense of "will." Volitional resolve is a commitment to act in certain ways, a determination to pursue certain ends, which, as Kant says in the *Groundwork*, implies the "straining of every means so far as they are in our control," even if the effort is partly or wholly unsuccessful. In an extended sense, we might

---


19By this, I mean ‘possible’ in other synthetic senses, such as causally possible, economically possible, socially possible, and so on; its being ‘volitionally possible’ is already presumed.

20Immanuel Kant, *Groundwork of the Metaphysic of Morals*, tr. H.J. Paton (Harper Torchbooks, 1964), Ak.3, p.62. Kant makes this remark in his discussion of the point that the will is "good through willing alone" rather than through its consequences or contingent material effectiveness: he wants to make clear that this thesis does not imply that a plan or idea of action abstractly considered can be good in itself, since if the agent (continued...)
even be able to say that we will certain lines of action (say, various political reforms) in the
counterfactual sense that we would commit ourselves to them if we could see a way to make
their effective pursuit within our power. But willing in this sense still has action as part of
its content.

However, this inevitable connection has led many philosophers and psychologists to try
to define will in terms of its relation to action, rather than in terms of its peculiar
motivational qualities. The tension between these divergent approaches to the "will" is
illustrated in the following passage from Yalom’s attempt to bring the concept of will back
into psychotherapy:

It is the mental agency that transforms awareness and knowledge into action, it is
the bridge between desire and act. It is the mental state that precedes action
(Aristotle). It is the mental "organ of the future"—just as memory is the organ of
the past (Arendt). It is the power of spontaneously beginning a series of successive
things (Kant). It is the seat of volition, the "responsible mover" within (Farber). It
is the "decisive factor in translating equilibrium into a process of change...an act
occurring between insight and action which is experienced as effort or
determination" (Wheelis) ...It is a force composed of both power and desire....To
this psychological construct, we assign the label, "will," and to its function,"willing."21

Many conflicting tendencies of our philosophical heritage are evident in this list. In the first
three sentences, volition is construed as the mental state which directly causes or
immediately precedes action, making preexisting motives effective. The next construal,
which is patently Kantian, qualifies this first analysis of will as an undetermined or
negatively free cause of action. But the quote from Wheelis focuses on some instructive
phenomenal qualities that characterize the experience of willing: namely, that it begins from
an equilibrial or relatively unmotivated state by a movement in which the agent feels that she

20(...continued)
is not making a sincere effort that at least tries to realize his purpose, he is not "willing" at all.
21Yalom, p.289-90.
is positively asserting herself, which suggests that volition is a *motivation-forming*, not merely action-determining, attitude.

Yet because we must consider conditions for moral responsibility and the imputability of actions—as well as other reasons\(^2\) we are interested in the senses in which our actions may be "free." This connection between action and freedom, which underlies the debates between different compatibilist and libertarian accounts of ‘voluntary’ action, has led many philosophers to conceive the will in terms of their preferred sense of freedom to choose our actions. The problem with this approach is that we have to have some grasp of what volition is to guide our inquiry into what kinds of freedom it might or might not have. When will is conceived in advance in the thin fashion, i.e. as the capacity to act ‘freely’ in one or another contentious sense, it gets separated again from any distinctive sort of motivational experience. Consider for example Yalom’s claim that "Motivation can influence but not replace the will; despite various motives, the individual still has the option of behaving or not behaving in a certain fashion."\(^3\) When will is simply equated with freedom in such a libertarian sense, volition is reduced to a pure act of consent that stands above all motivations, deciding to form an intention on the basis of some reason or desire, rather than itself constituting a distinctive kind of motive. This separation has led to a host of interrelated objections that the will is only an ‘arbitrary’ or ‘decisionistic’ power whose deliverances cannot themselves be rationally explained or substantively motivated. It is

\(^{2}\) In particular, the problems of theodicy, the origin of the intelligibility of non-actual possibilities (and hence modality in general), and the experience of time as a real A-series with an open future, may all require some robust concept of alternative-possibilities liberty as their ground of possibility.

\(^{3}\) Ibid, p.291. Yalom’s belief that volition must be something distinct from any kind of motivation arises in part from a limited conception of the kinds of motivation there can be: if motives are exhausted by desires, instinctive urges, and emotional states, as many writers assume, then of course volition must be something else. But there is another possibility that is being overlooked here: namely that the might be a distinctively volitional type of motivation itself, which (unlike the standard types of motivation) could never be ‘involuntary’ or unwanted by the agent.
unfortunate that such a direct equation of will with a liberty for alternate actions is now widely thought of as representing ‘the existentialist’ position. For philosophers in the existentialist tradition instead start with more substantive conceptions of volition as an intrapersonal phenomenon —conceptions which leave the issue of ‘freedom’ of the will as at least a conceptually separate question from the motivational nature of willing per se— despite the fact that they also hold some qualified libertarian conditions for moral responsibility. By contrast, the simplistic equation of the will with indeterministic freedom of action is more characteristic of the libertarian literature in contemporary analytic metaphysics. Professor Van Inwagen provides a prominent example:

I use the term ‘free will’ out of respect for tradition. My use of the term is not meant to imply that I think there is such a "faculty" as "the will." When I say of a man that he "has free will" I mean [only] that very often, if not always, when he has to choose between two or more mutually incompatible courses of action...each of these courses of action is such that he can, or is able to, or has it within his power to carry it out.

This definition is consonant with the thin formal notion of volition as simply the proximate mental cause of actions (usually bodily acts) or decision to act, assuming that Van Inwagen’s alternative "courses of action" each require something like an intention.

This contrasts with the ‘thick’ formal concept of the will as a distinctive kind of

---

24There is admittedly a strand in so-called ‘existentialist’ literature that does provoke this common attribution of a reductive conception of will to existentialists. This strand, which sometimes appears dominant in Sartre (though even in Being and Nothingness it does not totally exclude more substantial elements of volition) may ultimately derive from Fichte’s claim that since "nothing pertains to the I which it does not posit," its feelings of limitation must also be posited or self-given (a position Fichte shares with Reinhold). See §6 of Fichte’s summary, "The Major Points of the Wissenschaftslehre of 1789-1799," in Foundations of Transcendental Philosophy: (Wissenschaftslehre) Novo Methodo (1796/99), tr. & ed. Daniel Breazeale (Cornell University Press, 1992), p.69.

25The textual exegesis which would demonstrate this thesis is central to the second part of the overall project of which this dissertation forms the first part, as described in the prologue above. But see, again, my papers on Kierkegaard’s Either/Or and Concept of Anxiety.


27Otherwise Van Inwagen’s definition would allow will to any system that could behave different ways in a single circumstance, even if it lacked mental life or consciousness entirely, and so would be compatible with a libertarian behaviorism.
motivational capacity, which involves but cannot be reduced to deciding or intending to act in some way or under some description. Different scholastic conceptions of the will as a "faculty" were thick or substantive conceptions in this formal sense, but the notion of a "faculty" as a metaphysically distinct vital power is not necessarily required for a thick conception of volitions. What all such conceptions share is their understanding of the phenomenon to be explained as ‘will’ in the sense which is uppermost in prephilosophical usage or ‘ordinary language:’ namely, volition as personal resolve, or choice that is motivated by the agent’s assert self-commitment. As Yalom notes, the word "will" has rich connotations precisely because "it conveys determination and commitment," like a promise. Of course, the will’s motivational nature and its freedom are intimately interconnected questions: for example, if a philosopher argues that volitions are decisions with the composite motivational character of ‘rational appetite,’ this entails certain difficulties in maintaining a libertarian conception of the will’s freedom; other conceptions will lack these difficulties but entail other constraints on freedom.

Similarly, whether a philosopher approaches will as a thin or thick notion affects any later account of freedom in willing which she develops. In particular, if she develops a libertarian theory of the free will, it will affect her resources for providing an adequate answer to

---

28Hence thick conceptions of volition offer differing theories, explanations, or interpretations of will in the motivational sense, but they share the same formal concept of what it is they are attempting to explicate. This formal notion of the will as the capacity for a particular kind of motivation is not itself a "theory" of the will. Thus if this thick formal concept represents our ‘folk-psychological’ notion of volition, folk psychology does not on this account give us any theory or determinate conception of the will. It only proleptically demarcates or identifies the phenomena that such theories are supposed to explain. This function of setting formal concepts rather than theories, or determining the problematic which developed conceptions try to solve, is the usual role of ‘folk psychology’ —or, more accurately, of the complex hermeneutic ‘background’ of inherited ideas and assumed usages affected by past theoretical interpretations, for which "folk-thinking" is the shorthand in contemporary analytic philosophy.

29Yalom, p.291.

30This does not mean that a full conception of free will as rational appetite with a limited sphere of alternate possibilities is impossible, however: see my article, "Aquinas’s Teleological Libertarianism," in The Modern Schoolman, forthcoming.
traditional difficulties, such as reductios claiming to show that libertarian freedom terminates in arbitrariness. Note that only on a thick or motivationally substantive notion of the will could there ever be dispositions that were distinctively volitional, i.e. not mere dispositions to act or behave in certain ways, but dispositions to will. This leads to an important hypothesis which is a fundamental part of the motivation for this dissertation: if existentialism is to be a viable position on the nature of personhood, existentialists must develop a thick conception of the will which remains compatible with volitional liberty, because only on such an approach can there be a serious prospect of overcoming the challenge of the arbitrariness objection.\

However, in contemporary analytic philosophy, as in most psychological research on action, the thin concept of volition as first-order intentional action reigns supreme. Not only libertarians, but also their most ardent opponents, offer only conceptions or theories which attempt to spell out or explain the will in its thin formal sense. Within this thin paradigm there thus remains a wide variation of different conceptions in philosophical and psychological analysis of action. Some try defend fairly elaborate antireductionist theories of volition, such as Brian O’Shaughnessy’s in his massive two-volume study. At the other extreme but still within the same paradigm of the thin concept, some take the eliminativist line that because the conceptual vocabulary of intentional action is sufficient by itself to account for everything we need, we should recognize that our (thin) concept of "will" is empty and dispense with it altogether. Thus, following G.E.M. Anscombe, Daniel Dennett says:

31 See the Prologue for a discussion of the dissertation’s role in this larger project of providing a complete existentialist account of personhood.

32 Brian O’Shaughnessy, The Will, in two volumes (Cambridge University Press, 1980). O’Shaughnessy meticulously works through different possible accounts of the mental antecedents to physical or bodily action, but he does not treat the topic of autonomy in one’s motivations: for example, there is no discussion of Frankfurt’s hierarchical conception of volition in his study.
The account of intention that has been given includes no talk about volitions or willing. That is because, as Anscombe argues, the verb ‘to will’ is a hoax. There are no such things as acts of will or volition\textsuperscript{33}....The idea that willing is some sort of radiation generated by gritting the teeth and saying, ‘move, move, move,’ is hopeless. It arises, no doubt, from such experiences as lying in bed and saying to oneself ‘I must get up, I must get up, it’s late. On the count of three: one, two, three,’ until one finally gets up.... The fact of the matter —that sometimes the thoughts seem to help and sometimes not— suggest that thinking to oneself is merely an accompaniment or by-product of the actual business of determining action.\textsuperscript{34}

Instead, Dennett argues, the brain makes the decision one way or another unconsciously, according to whether the right "balance is tipped" in one’s neural activity, which need not be recognized verbally so that one seems to ‘will’ to get up.\textsuperscript{35} In this case, giving verbal stimulation to one’s own neural activity is exactly analogous to some third party giving such stimulation: "the success in both cases, of course, would depend on the relevance and abundance of the information produced," which need not be conscious.\textsuperscript{36} In sum,

The notion that must be avoided is that awareness is in any way a centre from which efficacious signals, volitions, or any sort of psychic radiation emanates.\textsuperscript{37}

Note that while philosophers like O’Shaughnessy and Dennett disagree radically about whether the psychic states which cause bodily actions are essentially conscious and have a discrete enough role in conscious mental life that the thin concept of ‘will’ finds any purchase, they nevertheless agree that "willing" is merely bringing about first-order acts as intended, something that many animals can do as well as human beings.

\footnotesize{\textsuperscript{33}At this point, Dennett quotes Anscombe’s complaint that there is no univocal sense in which I can ‘will’ my arm to move, but not will a matchbox to move —see Anscombe, \textit{Intention} (Oxford University Press, 1957), p.48-9. This sort of view is the main target of O’Shaughnessy’s work, which constitutes as imposing and comprehensive a refutation as one could imagine of this Anscombian/Dennettian position.


\textsuperscript{35}Ibid, p.172.

\textsuperscript{36}Ibid, p.173. Dennett equates conscious "awareness" here with verbal expression of the information which the brain is processing.

\textsuperscript{37}Ibid, p.173.}
I.1.2. ‘Volitionalism’

As it happens, there is now much empirical evidence against Dennett’s eliminativist conception of (thin) willing, and more robust versions of "volitionalism" are common in action theory. In his helpful summary of the current literature, Robert Audi surveys several "varieties of volitionalism," which means "a theory of the nature of action which gives a central place to one or another kind of willing." Several of these theories, such as Hugh McCann’s, portray volitions as mental "actions" that prompt outward performances, or as in Lawrence Davis’s theory, as "attempts or tryings," which are a special class of actions. Not only O’Shaughnessy, but also D.M. Armstrong and Raimo Tuomela "conceive volition as roughly equivalent to trying" which causally explains action. Others, such as Michael Zimmerman, have equated volition with a "kind of decision," while philosophers in the Millian tradition, including Wilfred Sellars and Myles Brand, have described volition as one type of "intention" which is the "proximate cause of action." It is clear that all these conceptions start with the thin concept of will as the mental state which is necessary or even sufficient to cause intentional bodily movements (or outward ‘acts’), thus turning the question of volition into a particular issue in theories of action which are usually meant to apply (at least in principle) not just to persons, but potentially to other animals as well. In

---

38See Laurence Miller, *Inner Natures: Brain, Self, and Personality* (St. Martin’s Press, 1990), Part II, ch.3: "The Impetus of Self: Action and Volition." Miller argues that it is portions of the frontal lobe which control "the overall regulation and evaluation of thought and behavior" (p.44). Damage to this part of the brain leaves more particular abilities intact, but destroys regulation of their sequences: "In such cases, the individual’s very capacity for autonomous volition seems to have been sucked right out of him....the impression is not unlike that of a robot with a faulty guidance system: The patient does little on his own, but once prompted, seems incapable of stopping what he’s already doing or shifting to something else without explicit outside guidance" (p.45).


40Audi, p.75.

41Audi, p.76.

42Audi, p.76.
the few cases where contemporary analytic treatments have tried to describe volition in terms of motive, they have generally reduced it to some type of desire: e.g. "For Alvin Goldman, occurrent (predominant) wanting is approximately equivalent to what has been commonly considered volition..."\(^{43}\)

As Audi explains, some "volitionalist" accounts recognize volitions as having a special kind of self-referential intentional content. For example, Carl Ginet holds that in exerting bodily movement, I do not simply will the exertion: rather, I will "my exerting...I will that my willing—this very willing of whose content we speak—cause the exertion."\(^{44}\) This is insightful, since it brings out the essential reference to the agent’s individuality in the content of volition, but it does not by itself recognize that willing is second-order act with a distinctive kind of motivational quality.\(^{45}\) What unites all these "volitionalist" accounts, as Audi points out, is just the conviction that volition is a class of mental events that "are, or are crucial in, producing action—if not its immediate causes, then its closes psychological causes."\(^{46}\) The point of such theories is to start from "the picture of actions as caused by such elements as the agent’s desires, beliefs, and decisions" and provide a "causal factor which genetically unifies actions in terms of a common kind of origin, even if not necessarily an ultimate origin, in the psychology of the agent."\(^{47}\) In this dominant contemporary outlook, will is interpreted relative to a set of problems in theories of action

\(^{43}\) Audi, p.77.


\(^{45}\) It points in the right direction since, as we will see, what distinguishes volition is the sense that the agent herself is the *initiator of the motivation* by which an object becomes a goal or end for her.

\(^{46}\) Audi, p.78. Yet Audi draws the wrong conclusion when he infers from the thinness of this agreement that "There may be no ordinary conception of volition, or willing, or even trying, which is being analyzed and used to understand action" (p.79). Rather, all these accounts are conceptions of the *thin formal concept* that remained after the historical degeneration of the notion of will in the empiricist tradition form Hobbes through William James.

\(^{47}\) Audi, p.79.
which already assume that motivation is defined by the limits of interaction between beliefs and desires—precisely the presuppositions which the phenomena underlying the thick concept of will should give us reason to challenge.

I.1.3. Voluntariness and Autonomy

The thin concept also remains dominant in domains outside metaphysical action theory. Historical accounts, such as Hans Oberdiek’s history of the will, tend to stress the idea that volition is used "to distinguish between intentional and unintentional action, a grave matter in many criminal law cases." For example,

Self-controlled action, as Aristotle explains it, serves as an important forerunner to our own notion of ‘voluntariness.’ For in Aristotle’s account, self-controlled (or voluntary) actions are those where an agent, acting with the appropriate degree of knowledge and without compulsion, seeks an end because of some kind of desire (orexis) originating within him.

Oberdiek sees the contributions subsequent to Aristotle primarily as giving us more adequate criteria for making this distinction, e.g. by adding the idea that the will is "a phenomenon of subjective, introspective consciousness, a mental event preceding action" (Descartes), and by adding the notion of intention to distinguish human actions from mere behaviors (Aquinas). The function of the will as a philosophical concept on Oberdiek’s reading thus agrees with today’s volitionalists:

A theory involving 'volitions promises not only to distinguish involuntary from voluntary behavior, but to explain how voluntary actions occur and what counts as

---

49 Ibid, p.466.
50 Ibid, p.466.
51 Ibid, p.469.
Similarly, in his own interesting account Oberdiek argues (following Davidson, Donagan, and others) that act-descriptions and predications of ‘is voluntary’ must be treated intentionally: "an agent acts voluntarily only when the act-description is one which the agent is aware describes his act." Although he distinguishes between "free" and "voluntary" actions, so that coerced actions may be voluntary yet unfree, Oberdiek’s account still equates volition with voluntary action: e.g. a wink is willed, a blink is not. Though like other volitionalists his account is antireductivist and convincingly responds to reductionists like Ryle, it still treats the will only in terms of first-order agency.

Since many theories of "autonomy" employ a notion of "freedom" distinct voluntariness—as in Oberdiek’s account— one might assume that autonomy will involve the substantive or thick notion of volition lacking in proposed conditions for voluntary first-order agency. This is true for some accounts of autonomy, such as Frankfurt’s, which will be analyzed in detail in Chapter IV. But there are many purported accounts of autonomy which still ignore second-order agency, and define autonomy in terms of voluntary action plus other external conditions. For example, Gert and Duggan distinguish three senses of voluntariness:

In ordinary discourse the term "voluntary" as applied to actions has a number of senses which though related are, nonetheless, distinct. Of these, three are central: (1) a voluntary act is simply an intentional (as opposed to accidental) act; (2) a voluntary act is a free (as opposed to constrained) act; (3) a voluntary act is an intentional act done by an agent who has what we have described as the ability to

52 Ibid, p.470.

53 Ibid, p.485. He also notes that "Descriptions under which an act is voluntary are not always descriptions under which they are intentional" (p.484), since an act is voluntary under a description of consequences that are foreseen, but foreseen consequences may not always be intended. This suggests that intention, unlike voluntariness, makes implicit reference to the motives for which an act is done: these motives pick out a subset of the descriptions under which an act is voluntary. But Oberdiek does not note that such a distinction would not apply to animals capable of voluntary acts but lacking second-order agency, who do not decide to form intentions to act for specific reasons.

54 Ibid, p.483.
will (volitional ability) to do that kind of action.  

The first two conditions are roughly equivalent to voluntariness in Oberdiek’s sense; the third add to this a complex condition requiring that the agent can recognize "coercive" and "non-coercive incentives" for doing a particular act X, and would usually respond to them in appropriate fashion, e.g. by not doing X when coercive incentives for not doing it are present, by sometimes doing it when non-coercive incentives for doing it are present, etc.

There are many kinds of incentive: "moral, prudential, patriotic," monetary, etc., and in order to act freely, an agent need only be able to have beliefs that there are incentives of the many kinds normal people recognize, and respond to them such that he "almost always" acts in accordance with coercive incentives, and sometimes acts in accordance with non-coercive incentives. For example, a son who obsessively visits his mother’s grave due to guilt feelings either cannot "believe that he should not visit her grave, or if he can come to believe this it does not affect his actions." Gert and Duggan suggest that those who cannot recognize common incentives as they require "are in that respect like delusional psychotics," while those who recognize but do not rationally respond to incentives "are in that respect like compulsive or phobic neurotics."

Like other treatments (e.g. John Fischer’s) this analysis assumes that the idea of autonomy is meant primarily to embody standards of normal reason-responsiveness, and so Gert and Duggan build into it an objective (if rather unspecified) range of types of reasons or "incentives" that autonomous agents must be able to recognize and respond to appropriately.

---

57 Ibid, p.211.
58 Ibid, p.213.
59 Ibid, p.211.
They gloss over the problem that what such reasons or incentives are will in part be socially
determined. More importantly for my purposes, they entirely ignore the question of whether
autonomy requires any ability to control one’s incentives oneself, or any distinctive kind of
motivation (which may hardly be ‘incentival’ in form). Thus they still conceive volition
entirely in terms of first-order agency: "we define ‘willing’ as doing intentionally or trying to
do." Thus autonomous willing is merely intentional action that is sufficiently reasons-
responsive.

Other accounts of autonomy do conceive the will as second-order agency, but like Pink,
do not recognize in such agency any motivational capacity distinct from desires or practical
reasoning about the good. For example, Stampe and Gibson define "the will" as "the
intention-forming powers of the mind, or traditionally, ‘the power of choice’," and define
"'decision’ or ‘choice’ to mean just a mental event or act of mind in which an agent came to
have the intention to do something A, where that event or act is one of such a kind that the
agent’s reasons for doing A might have been the cause of that intention." This is a weaker
condition than Pink’s reason-applying conception of volitional decision, but it rules out
intentions installed by neurologists, or brought about by hypnotic suggestion, for example.
They argue convincingly that "internal barriers" may prevent a person from deciding or
forming an intention to act freely, and thus prevent them from acting autonomously: for
example, a agoraphobic woman does not freely remain in her room. But they explain this
by saying that her decision to remain in the room is unfree because the lacks "the condition
of the will that makes it possible for an agent to make the rational decision about what to

---

60Ibid, p.209, note 5.
61Dennis Stampe and Martha Gibson, "Of One’s Own Free Will," Philosophy and Phenomenological
Research, 52.3 (September 1992): 529-...., p.530.
63Ibid, pp.529-530.
the rational action being what would be best "relative to" the desires or some agent and the evidence he or she actually possesses. (Thus, to be able to make "the rational decision" is, on one standard view, to be able to make that decision the implementation of which would tend to maximize the satisfaction of one’s desires, if the relevant beliefs were true. 64

Stampe and Gibson anticipate the objection that in this sense, the agoraphobic woman does decide rationally, since she satisfies her strongest desire (i.e. to avoid open spaces), though at the cost of frustrating other desires, like the urge to see her friend. 65 But they suggest that what makes her will unfree is rather that counterfactually, even if "in her own estimation" going out were worth the risk of a panic attack, she still would not decide to do what would then be the subjectively rational thing for her to do: her will cannot track rationality in alternate possible variations of her circumstance. Similarly, what makes us unfree in a holdup is that, even though we act rationally in handing over our wallet to the gunman, were we to believe that the rational thing was to defy him, we would be too scared to do it. 66

Stampe and Gibson compare this notion of free will to a "free-working mechanism: like a weathervane, which is not stuck in one position, but rather can point in the direction of the wind, whatever direction the wind may blow." Although a stuck weathervane may point in the right direction if by accident the wind is flowing in that direction today, it is not free.

64Ibid, pp.532-3.

65Granted, the agoraphobic may not deliberate about this, since she is seized with an overwhelming fear at the bare thought of opening the door and cannot contemplate it. But Stampe and Gibson allow that "a decision need not be an act of deliberation, and it need not be conscious" (p. 530, note 3).

66Ibid, p.532. In general, Stampe and Gibson say, freedom of will requires the ability "to make the rational decision even if the circumstances were relevantly different from the way they actually are—e.g. just different enough that the rational decision would be something other than the one that is rational in the actual case" (p.532). But the overlook the problem that one of the most important ways in which the circumstances can vary is for the agent’s desires to change, or for the desires controlling intention-formation to alter. And since the will is often aimed precisely at bringing about such changes, its freedom can hardly be defined according to whether the decisions it would form in a range of possible circumstances would be rational relative to the strengths of desires given in these different scenarios.
since it does not track the wind.\footnote{Ibid, p.533.}

In response, we might doubt that a robot which perfectly decided the best action in the circumstances relative to the goal of fulfilling its programming could have anything like "a will." But aside from this intuitive problem, Stampe and Gibson’s conception of autonomy offers an implausibly indirect account of the feeling of coercion in such situations. Even if I thought I could defy the gunman had I reason enough to do so (e.g. if my wallet contained the only copy of a formula for a cure to AIDS) the fear on which I act in the actual sequence where it is not outweighed is not a motive with which I identify, in Frankfurt’s sense.\footnote{Science-fiction fans may consider the example of Angus, a cyborg in Stephen Donaldson’s novels A Dark and Hungry God Arises and Chaos and Order (Bantam Doubleday, 1994). Angus’s mind has been "welded" to a minicomputer in his back, which controls a whole net of minielectrodes in his brain, and these in turn control his every bodily action in accordance with pre-written instructions. Angus does not act like a unconscious machine: he forms intentions and even whole plans of action, which are always subjectively rational relative to his enforced desire to achieve his programmed goals. But he very clearly does not form these intentions freely: his will helplessly rebels against the continual coercion, leaving Angus in a state of neverending torture. The second-order agency that is Angus is powerless before the non-agent machine that forms his intentions.}
The hypothesis that will involves a motivational capacity distinct from ordinary preferences and desires for our well-being allows for a more direct explanation of the agoraphobic’s unfreedom as well: she may resolve to leave her home with the plan of meeting a cherished friend downtown, even though she thinks this is irrational given her condition (the resolve thus expresses a kind of defiance or alienation of the emotion on which she still reluctantly grants that it is most rational to act). But she cannot form a real intention open the door, because the intentions guiding her first-order acts are formed by a strong desire for enclosure that overrides the alternate motivational force of her resolve. Unlike Stampe and Gibson’s, this more direct explanation does not equate volition with intention-control on the motive of subjective rationality, defined in terms of existing desires and beliefs.

\footnote{This notion of identification will be investigated in detail in Chapter IV.}
I.2. The Heritage of the Thin Concept of Will

The thin formal concept of will has a long history. In an instructive essay, Charles Kahn traces four "different families of philosophical theories of the will" which each start from different preliminary concepts of the phenomena which count as ‘volitional’ and need to be explained. The first is the "theological" notion of will in Christian philosophy, in which the thick concept developed from biblical sources. The second is source of the thin concept:

The post-Cartesian notion of will is the one most familiar to philosophers...in the English-speaking world. This essentially involves the notion of volition as an inner, mental event or act of consciousness which is the cause, accompaniment, or necessary condition for any outer action, that is, for any voluntary movement.

Note how well this description corresponds to Audi’s summary of the central core of contemporary ‘volitionalist’ theories of action. As Kahn notes, such "theories of volition from Hume to William James" emphasize the Cartesian mind/body distinction, focusing on volition as a "mental cause (or antecedent) of a physical act."

I.2.2. Hobbes, Hume, and Others

As we saw above, Hobbes’s definition of will launches this "post-Cartesian" theory, eliminating the Platonism which remained in Descartes’s own account of the will. As

---


71 Kahn’s third tradition "begins with Kant’s notion of the will as self-legislation," and leads to a conception of volition as expressing the noumenal reality of the self, beyond the reach of theoretical cognition (p.236). Yet, as we will see, Kant’s notion of motivation draws on the thick concept of volition that originates in Kahn’s theological tradition. Moreover, Kahn’s fourth notion —free will vs. determinism— is not really a family of theories or a tradition, but a problem whose meaning depends on which of the other three notions of volition we have started with.

72 Kahn, p.235.

73 Kahn, p.235-6; though a mental act such as remembering could also presumably be ‘willed’ in this thin sense if it had a mental antecedent such as ‘intending to remember.’
Oberdiek comments in his history of the will, Hobbes’s doctrine that deliberation "amounts to nothing more than a tiny war of desires and aversions...would seem to eliminate the will altogether," at least in the sense of an independent motivational faculty.\textsuperscript{74} Hobbes insists that freedom is simply the ability to do what one desires without external constraint, and that our psychology involves no "second-order agency of the will" which could stand above states like desires and inclinations in order shape from them persisting motives that will support intended plans of actions into the future. As Pink argues, this denial of second-order agency has proved enormously influential since. In particular, it has marked Anglo-Saxon theory, from Hume through to Ramsey, the pioneer of modern rational choice theory, and Ryle....And this Anglo-Saxon tradition of disbelief or at least doubt in second-order agency still has plenty of supporters. Daniel Dennett, for example, has wondered whether there really is an agency of decision-making which explains our actions...And in his moral psychology, as we shall soon see, Bernard Williams denies that action arises out of second-order agency of the will.\textsuperscript{75}

In this respect, as Pink sees, the doctrinal "internalist" conception of motivation in analytic philosophy remains emotivist or broadly Hobbesian. Bond summarizes this doctrine as follows:

\begin{quote}
It is the favoured view among professional philosophers at present, and I include Bernard Williams, J. L. Mackie, Richard Taylor, Gilbert Harman, and Rodger Beehler, that nothing can be a [practical] reason for a person unless it somehow relates to his or her present or actual set of desires, wants, or motivational propensities....This is the pure doctrine. An agent can only deliberate on the basis of the wants she already had....Deliberation is of means alone.\textsuperscript{76}
\end{quote}

However, there is a long path between Hobbes’s revolution and the reductionist theories of motivation that dominate 20th century analytic philosophy. Although I cannot retrace that path in detail here, looking briefly at a few points along the way may help clarify how the

\textsuperscript{74}Oberdiek, "The Will," p.471.
\textsuperscript{75}Pink, p.26-27.
\textsuperscript{76}E. J. Bond, \textit{Reason and Value} (Cambridge University Press, 1983), p.3. Bond notes that Williams allows deliberation a role in \textit{indirectly} changing one’s "internal set" of ends, but "all such changes are nevertheless related to the more efficient satisfaction of desires one already has" (p.4).
thin concept became an ideology.

Following Hobbes, Locke attacked the notion that the mind contains distinct "faculties" as implying that each is a "distinct agent" within us, and defined the will as an unanalyzable power of bringing about change, and Hume followed him by defining the will as "the internal impression we feel and are conscious of, when we knowingly give rises to any new motion of our body, or new perception of our mind." Thus the will is merely the subjective impression of an intentional initiation of first-order action. Hobbes follows this by arguing that the "different stations in life" or natural classes and "uniformity of human actions" within nations and individuals show that psychology is run by the same kind of regular laws as physical mechanics. Since in ordinary life we constantly rely others' actions being predictable and rooted in stable dispositions, "whoever reasons after this manner, does ipso facto believe the actions of the will to arise from necessity." And even more influentially, Hume argued, against the prejudice that reason should rule the passions, "that reason alone can never be a motive of any action of the will" and "can never oppose passion in the direction of the will." Thus the will was not only generally reduced to a power of first-order intentional action, but more specifically conceived as the economy of desires and

79Ibid, p.402-3. The argument here is a direct precursor of the kind we now find at the foundation of personality type theories, which begin by holding that statistical consistencies in behavior patterns cannot be denied and reveal the existence of traits. Similarly, Hume appeals to the idea of stock character types in saying: "No union can be more certain, than that of some actions with some motives and characters." Indeed, irregularity in behavior is a sign of madness (Ibid, p.404). Actions are a function of "motives, temper, and situation" (Ibid).
80Ibid, p.405.
81Ibid, §3, p.413.
aversions arising from expectations of pleasure and pain.\footnote{Ibid, p.414: "Tis from the prospect of pleasure or pain that the aversion or propensity arises toward any object."} Although Hume offered a more nuanced account of our natural passions than Hobbes and Locke (including "calm desires and tendencies," such as social dispositions\footnote{Ibid, p.417.}), he sets the stage for later utilitarian models of motivation\footnote{See Oberdiek, "The Will," where he described J. S. Mill’s Humean notion of the will as "that moment in a series of associations of ideas terminating in action," or the immediate mental antecedent to action (p.474).} and the notion of the human agent as a simple egoist—a utility or preference-satisfaction maximizer—that is canonical throughout modern economist theory.

In the same period, the Cambridge Platonists and their followers, such as Jonathan Edwards, agreed with Locke in rejecting faculty psychology and freedom as simple alternate-possibilities liberty, but held to a more rationalist theory of the will as cognitively motivated by judgments about the value of ends. In his \textit{Freedom of the Will}, Edwards defends the idea of "moral necessity," a psychological determinism in which "the will always follows the last dictate of the understanding."\footnote{John E. Smith, \textit{Jonathan Edwards: Puritan, Preacher, Philosopher} (University of Notre Dame Press, 1992), p.66.} Our will is determined by the strongest motive, which is equivalent to "the mind’s apprehension of the greatest apparent good," although the particular way our apprehension works develops like a habit over time (and is influenced by grace).\footnote{Smith, p.64-65.} Thus while the will is cognitive, it is teleologically determined: as Oberdiek says, for Edwards "the apparent good draws us towards it like a powerful magnet draws iron filings."\footnote{Oberdiek, "The Will," p.472-473.}

In his manuscript on free will, Ralph Cudworth similarly argued against the voluntarist idea of a "blind will" that "still remaineth as free, and indifferent to do or not this or that, as
if the understanding had given no judgment at all in the case, and doth at last fortuitously determine itself without respect to the same either way."\(^{88}\) Like Edwards after him, Cudworth stressed that will and understanding are not separate faculties, since it is the "man or soul" —the single agent or self— that does both,\(^ {89}\) and that the root of our motivation, or the essential propensity of our psychological nature, is the "constant, restless, uninterrupted desire or love of the good as such, or happiness" which means that we always will our apparent good.\(^ {90}\) But unlike Edwards, Cudworth also objects to the simple rationalist position that "the blind faculty of will always necessarily follows the last practical judgment of the understanding" about the greatest good,\(^ {91}\) since (as Aquinas held) the "good" we essentially desire is so general and hidden that our desire can be "diversely dispensed out, and placed upon different objects, more or less."\(^ {92}\) Yet Cudworth denies that this means the will can just choose indifferently between apparent goods, and again like Edwards, holds that the "hegemonic, or ruling principle in a man" is dispositional, and its understanding or judgment of the apparent good is shaped by prior choices (p.177). But unlike Edwards, for Cudworth this means that "this hegemonicon" is a "self-forming and self-framing power," itself activating man’s innate natural desire for the good, which through its own "purposes and resolutions...designs and active endeavours" alters its own dispositional tendency to judge or apprehend that good more in one way than in others (p.178).\(^ {93}\) This clearly seems


\(^{89}\)Ibid, p.171.

\(^{90}\)Ibid, p.173.

\(^{91}\)Ibid, p.168.

\(^{92}\)Ibid, p.174.

\(^{93}\)In this respect, it seems to me that Cudworth’s view tends away from Edwards’ determinism towards the existentialist conception (developed in Schelling and Kierkegaard) of free will as endowed with liberty, but (continued...
to imply a conception of the will as a second-order agency, contrary to Locke.

But these more subtle views about motivation did not win out against the doctrine of Hobbes and Hume that "every human act may be explained as the causal outcome of one’s desires and beliefs at the time of action." And it is not clear that Edwards and Cudworth’s Platonic conception of motivation as grounded in the desire for the good is logically incompatible with Hume’s position that it is not "contrary to reason to prefer even my own acknowledg’d lesser good to my greater." For Edwards and Cudworth must say that such a desire is unnatural, that it is desire not reason which is malfunctioning in Hume’s example. They agree with Hume that motivation terminates in beliefs about our apparent good and in consequent desires, although they give final control to the "general appetite to good, and aversion to evil," which Hume treats as just one motive among others in his list of "calm desires."

On the continent, the development of ideas on volition following Descartes is a complex story in which (especially following Kant) the thin concept is not quite so dominant. Yet, although I can barely touch on the tradition leading from rationalism to German idealism, it is worth pointing out that many writers in this lineage still take for granted the thin concept of willing. To take just one example, in his prize essay, Schopenhauer distinguishes between "freedom of action," which is the ability to behave in accordance with an already-given volition, and "freedom of willing." But even the latter only means for him the

\[\text{...continued}\]

working on itself as a substantive dispositional character whose freedom is therefore never arbitrary indifference.


\[96\] Ibid, p.417.

question of whether the will is free to form or not form an intention based on motivating objects presented by cognition. Thus his question is the same as Leibniz’s: "does the entrance of a motive into the consciousness necessarily bring about the volition, or does the will retain complete freedom either to will or not to will."\(^{98}\) Note that this question has nothing to do with the original formation of motives themselves. As his translator comments, for Schopenhauer, "the free will reveals itself in its tendency to respond to motives....the extent of its freedom will be proportionate to the scope of the motives to which it may respond." Thus freedom is ultimately traceable to the enlargement of the intellect’s cognitive powers.\(^{99}\) In this sense, Schopenhauer anticipated James’s approach.

### I.2.3. James

A century after Hume, William James gave a new account of the will\(^{100}\) that helped establish the dominance of the thin concept. From the Aristotelian premise that will concerns actions that are in our power, James first concludes, as if anticipating O’Shaughnessy, that "the only direct outward effects of the will are bodily movements."\(^{101}\) Voluntary control of movements is attained only after we have learned of these movements by experiencing them as involuntary behavior: "Reflex, instinctive, and emotional movements are all primary performances,"\(^{102}\) which build up the "kinaesthetic impressions"

\(^{97}\)(...continued)
dread freedom.


\(^{99}\)Ibid, p.xvi-xvii.


\(^{101}\)James, p.486.

\(^{102}\)James, p.487.
necessary for us to control our limbs through the body-image.\textsuperscript{103} In cases where the body-image is disrupted, such ‘volitional’ control over our movements is impaired.\textsuperscript{104} He then argues at length, in response to Bain, Wundt, and Helmholtz, that no "feeling of innervation" or "outgoing discharge" is necessary as an antecedent to action in voluntary bodily movements.\textsuperscript{105} Like Hobbes, James defends the view that "all our thoughts of movement [are] of sensational constitution," but argues that this still leaves room for spontaneity or "scope for our inward initiative to be shown" in focusing on some sensations rather than others.\textsuperscript{106} This link to sensation, which supplies our understanding of different movements and their ends, is needed to explain what that ‘idea of a movement’ is which must precede it in order that it be voluntary. It is not the thought of the innervation which the movement requires. It is the anticipation of the movement’s sensible effects, resident and remote, and sometimes very remote indeed. Such anticipations, to say the least, determine what our movements shall be.\textsuperscript{107}

James’s entire account of the will, then, amounts to spelling out in much more detail than Hobbes could (with the benefit of experimental results in 19th century psychology) the mechanisms by which states of "desire" are represented and can account for bodily movements. No notion of constituting motives by decisions, or sustaining motivation by committing oneself to care about some person, issue, or thing, ever enters into James’s analysis.

The nearest James comes to giving "decision" any role is in cases where ‘tie-breaking’ is

\textsuperscript{103}Compare O’Shaughnessy, \textit{The Will, Volume I}, pp.248-250.

\textsuperscript{104}James, pp.489-493.

\textsuperscript{105}James, p.493-516. At the end of all this, James concludes by saying "On the whole, then, it seems as probable as anything can be, that feelings of innervation do not exist" (p.516). The issues at stake in this debate are related to those which concern the contemporary theories Audi calls "volitionalist" theories, but have little relevance for my analysis.

\textsuperscript{106}James, p.518.

\textsuperscript{107}James, p.521.
needed between various sensual representations pulling on the agent. Sometimes "the bare idea of a movement’s sensible effects [is] its sufficient mental cue" and sometimes an "additional mental antecedent, in the shape of a fiat, decision, consent, [or] volitional mandate" is required.\textsuperscript{108} The former is voluntary "ideo-motor action," which makes up the "habitual goings and comings and rearrangements of ourselves which fill every hour of the day,"\textsuperscript{109} while the latter is a special case. James believes that "every representation of a movement awakens in some degree the actual movement which is its object," or produces a \textit{tendency} to move (however weak), and only "an antagonistic representation" can keep it from automatically producing ideo-motor action. This is so because "consciousness is \textit{in its very nature impulsive}," or movement-generating.\textsuperscript{110} Only in such cases, where there are competing tendencies, is "the express fiat, or act of mental consent to the movement" needed. What it does is to "neutraliz[e]...the antagonistic and inhibitory idea," letting the other representation take effect. Thus decision does not \textit{evaluate} motivating representations of sensible effects in any way distinct from that in which such representations themselves motivate us: it simply adds one more to the pile, releasing the "inner spring" that translates into "motor discharge."\textsuperscript{111} Hence, just as Hobbes said, the process of deliberation is simply a series or ‘economy’ of thoughts about action that are in tension, or mental representations of opposed objects of action:

\begin{quote}
The result is that peculiar feeling of inward unrest known are \textit{indecision}. As long as it lasts, with the various objects before the attention, we are said to \textit{deliberate};
\end{quote}

\begin{flushright}
\textsuperscript{108}James, p.522.
\end{flushright}

\begin{flushright}
\textsuperscript{109}James, p.523. Compare O'Shaughnessy on what he calls "sub-intentional actions:" \textit{The Will, Volume II}, ch. 10.
\end{flushright}

\begin{flushright}
\textsuperscript{110}James, p.526. As he says, "Every pulse of feeling which we have is the correlate of some neural activity that is already on its way to instigate a movement. Our sensations and thoughts are but cross-sections, as it were, of currents whose essential consequence is motion..." (p.526). This refers of course to James’s famous dynamistic theory of consciousness.
\end{flushright}

\begin{flushright}
\textsuperscript{111}James, p.527.
\end{flushright}
and when finally the original suggestion either prevails and makes the movement take place, or gets definitively quenched by its antagonists, we are said to decide, or to utter our voluntary fiat, in favor of one or the other course.\footnote{James, p.528.}

The act of fiat or consent is turns out to be the "last appetite" again.\footnote{Admittedly, James goes on to give a much more complex analysis of such deliberation, distinguishing five types of decision and different experiences of "will" that go along with them, but they all adhere to his basic Hobbesian paradigm.} Thus Oberdiek gives too much credit to James in suggesting that his notion of fiat shows that "will is not simply the last appetite or aversion adhering to action," but rather a way in which "We organize and systematize our diverse perceptions, aspirations, emotions, and beliefs through deliberation into a choice."\footnote{Oberdiek, "The Will," p.477.} For the deliberation preceding the fiat is just a more complex version of Hobbesian deliberation: "The process of deliberation contains an endless degree of complication. At every moment of it our consciousness is of an extremely complex object, namely the existence of the whole set of motives and their conflict."\footnote{James, p.528.} The outcome is a matter of "the oscillations of our attention" and the "associative flow of our ideas"\footnote{Ibid, p.529.} and not of any agency standing above these. So Oberdiek is right that James gives a criterion for voluntary actions as those "which must be performed attentively and which, in doing, we experience a feeling of resolve, effort, or fiat."\footnote{Oberdiek, "The Will," p.478.} But this does little more than fill out the type of "feeling" which Hume detected in first-order intentional actions.

\section*{1.2.4. Wittgenstein and O'Shaughnessy}

This Jamesian model set the tone for subsequent treatments of volition in both psychology and philosophy in the twentieth-century. For example, Ludwig Wittgenstein,
who influenced Anscombe, follows James in holding that "the prototype of the act of volition is the experience of muscular effort." Though there are apparently more ‘active’ cases, such as when "I deliberate whether to lift a certain heavyish weight, decide to do it, [...] then apply my force to it," this is not "voluntary" in any special sense that less reflective acts like casually scratching an itch or spontaneously getting out of bed are not: the voluntary or "volitional" is merely a family-resemblance concept. In general, however, "will" is just constituted by voluntary or intentional actions such as guided bodily movements or mental ‘acts’ like imagining. Thus Wittgenstein denies that willing is a separate action of deciding or motivating our actions, distinct from such ‘first-order’ intentional acts themselves:

In the sense in which I can ever bring anything about (such as a stomach-ache through over-eating) I can also bring about an act of willing. In this sense I bring about the act of willing to swim by jumping into the water. Doubtless I was trying to say: I can’t will willing; that is, it makes no sense to speak of willing willing. ‘Willing’ is not the name of an action; and so not the name of any voluntary action either.

This argument is the fountainhead of what Pink calls the "Pro-Attitude" interpretation (adopted by Anscombe, Davidson, Bratman, and others), which rejects the idea that volition is itself a kind of agency. If willing was an action, Wittgenstein thinks, it would be like trying to perform a first-order intentional act. But an act such as raising my arm is not experienced like trying to manipulate an instrument; there is no sense of effort other than

---

119Ibid, #11, p.150.
120Ibid, p.151-152.
122Ibid, I, #613, p.159.
that involved in actually tightening the muscles that move the arm. Thus "the sensation of innervation which is supposed to constitute the consciousness of the act of will" is merely that involved in intending any mental cognition or physical movement under our control. The subjective experience we call ‘volition’ cannot be anything more than kind of tension we feel in straining to lift a heavy weight, since there is no prior, distinct sense of trying to make such an effort: our outward behavior and its correlated kinaesthetic sensation are all we find when we introspect such supposed paradigm cases of ‘willing.’ Wittgenstein’s denial that we can ‘will willing’ is thus symptomatic of the attitude that volition is merely the subjective experience of attempting to do something already intended: on this assumption, will as a distinct or ‘higher-order’ agency in Pink’s sense could only be experienced as a trying to try, which must be either redundant or merely a case of instrumental self-manipulation. Will is not, as we imagine (due to deceiving linguistic analogies), a pure unmoved moving power or "motor which has no inertia in itself to overcome;" rather, since willing is not an act we can try to perform, the frustratable vs. unfrustratable distinction does

---

123Ibid, I, #627-28, p.162e. Of course, we can try to move our arm and discover that it is strapped down. But if I read him right, Wittgenstein means that there is no separate act of initiating the arm-movement which will either succeed if physically unimpeded or fail if the arm is held: for if there were, than this separate act would also be a trying, that could succeed or fail. But we do not experience this in voluntary acts like intentional bodily movements. I thus perceive no deep tension between Wittgenstein’s view and a theory like O'Shaughnessy’s, which simply unpacks the experience of intentional first-order action to show that it involves a sense of trying to perform the intended act. Wittgenstein is out to deny that we try to form first-order intentions themselves, as Pink maintains.


125Ibid, #766, p.137e. Actually, Wittgenstein often uses "trying" in the more specific sense of making a focused effort against resistance, since he says that in ordinary cases, "when I walk, that doesn’t mean that I try to walk and it succeeds" (Remarks on the Philosophy of Psychology, I, #51, p.13); this explains why he says that although an experience of effort is the paradigm, it is not essential to volition, since it is lacking in voluntary acts like handwriting or moving the eye to look at an object (The Blue and Brown Books, Brown Book, II, #13, p.152-3). Moreover, it is crucial to Wittgenstein’s view that the kinaesthetic sensation we feel in, say moving our hand while drawing, is not something we anticipate and decide to bring about: at the most basic level in voluntary action, there must be motions or changes that are part of the ‘matter’ of the action, but are not themselves specifically selected in advance by deliberate choice (Ibid, #13, p.154). These reflections, like O'Shaughnessy’s work which builds on them, shed light on the structure of first-order intentional acting (especially in typical outward bodily cases) but they tell us nothing about will in the thick or agentive sense.
not apply to it at all.\textsuperscript{126}

Citing this very passage, O'Shaughnessy similarly argues that willing itself cannot be willed. He says that $\Phi$ is "the willing of some $\varphi$" if $\Phi$ is "the act of bringing about $\varphi$," where $\varphi$ is, for example, a "limb movement" like "the event of an arm rise."\textsuperscript{127} Thus construed as the carrying out of a pre-existing intention, a further act $\theta$ that wills or brings about $\Phi$ would make no sense. For example, if nerving or "getting myself" to shoot someone ($\theta$) does not refer to some kind of imaginative self-manipulation (like promising myself rewards for doing it), then it cannot be distinct from the act of shooting him ($\Phi$), which is "the making happen of a finger movement" on the trigger ($\varphi$).\textsuperscript{128}

The problem with these arguments, however, is their assumption that acts of will as a ‘higher agency’ would again be purely executive in character, carrying out a predetermined intention (in this case, an intention to form first-order intentions to act). As Pink notes, the "Pro Attitude model" which denies higher agency is moved largely by the notion that nothing which was not a reason or justification for doing $A$ could still be a reason for deciding to do $A$, and so the latter is superfluous.\textsuperscript{129} Yet as Pink notes, our control over our decisions or intention-formation is "direct" rather than merely instrumental:

The agency which gives us decision and intention control is agency which constitutes the taking of a particular decision to act, rather than agency which causes that decision to be taken.\textsuperscript{130}

Moreover, if volitional agency helps produce the very motivation on which it acts in forming intentions and plans of future action (as I will argue in Chapter II), then its distinctness from first-order action will be clear.

\textsuperscript{127}O'Shaughnessy, \textit{The Will}, Vol. I, p.29.
\textsuperscript{128}Ibid, pp.32-34.
\textsuperscript{129}Pink, \textit{The Psychology of Freedom}, p.9.
\textsuperscript{130}Pink, \textit{The Psychology of Freedom}, p.17, my italics for emphasis.
I.2.5. Other Recent Treatments of the Will on the Thin-Concept Paradigm

The lineage of the thin concept of will continues from Wittgenstein to Anscombe, Ryle, Dennett, Davidson, and their critics in the ‘volitionalist’ school, becoming part of the basic pedagogy of analytic philosophy of action by the turn of this century. But in this lineage, there is thus no serious or penetrating argument against the alternative thick concept of volition as motivational agency. Rather, the thin notion of volition became part of the horizon in the dominant outlook, obscuring the very existence of any alternative here. A good measure of this is the fact that in the 1960 Encyclopedia of Philosophy, there was no entry on "Will," and the entry on "volition" was devoted entirely to theories of what makes something a voluntary or intentional action.

Many writers today start with the thin-concept as their implicit paradigm. For example, in a recent paper, N.M.L. Nathan sketches a new conception of personal identity over time based on the idea that any two active volitions (e.g. "attending, deciding or putting forth an effort") share a "determinable" quality, but no two simultaneous volitions can share the same determinate instantiation of this quality —and similarly for all other experiences that instead have a "passive volitional nature," or a sense of being affected.\(^{131}\) Nathan believes this quality may provide what is needed for an "internalizing mental biography," that is one whose concepts "are describable in terms of a relation in the set of mental items which can be specified without reference to anything of a non-mental kind."\(^{132}\) Manifestly, such a project presupposes not only that all our experiences have just one of these two basic qualities, but also that we can isolate what counts as "volitional" adequately by thinking


\(^{132}\)Ibid, p.82.
about a range of mental and bodily actions without any attention to the formation of motivation and the structure of character. Thus the moral psychology employed in metaphysical analyses becomes so thin relative to the moral psychology developed in analyses of moral selfhood that prescind from metaphysical analysis.

This tendency to take the thin notion for granted results in cases like Allan Harkavy’s well-intentioned but philosophically naive attempt to postulate a "physical basis" for the human will in quantum-mechanical information coded in brain states. Harkavy says that in our ordinary conception, "'Will’ as distinct from consciousness results in motion," and yet must derive from knowledge, whose physical representation is information. Therefore, "If information in material systems could somehow result in non-random motion, we would have a physical basis for human will." Behind this inference lies a notion of the will so thin that it requires nothing more than informationally guided bodily motion, without any motivational phenomenology at all. That Harkavy could so easily make this assumption testifies to the dominance of the empiricist paradigm which has so radically reduced volition to a shadow of its former significance. As a result of this long tradition, in contemporary philosophy of mind today, so much effort is now spent on whether such basic concepts of ‘folk psychology’ as belief and desire can be defended as real mental states distinct from the neural states on which they supervene, that current debates can never reach the point of considering if there is a further sort of motivation which is not merely cognitive or appetitive.

I.3. The Thick Concept of Will and ‘Middle Soul’ Motivation

---

133 Allan A. Harkavy, Human Will: The Search for its Physical Basis (Peter Lang, 1995).
134 Ibid, p.11.
The thick concept of will can be traced, as Pink suggests, back through both Aquinas and the Stoics, who both conceived the will in terms of second-order agency. The Stoics distinguished human from animal agency by "the human capacity for sunkatathesis, or making assent to propositions about how one should act." But like Aquinas after them, the Stoics conceived such intention-forming acts of 'consent' as "an agency of practical judgment," which is an integral part of our species-essence as 'rational animal.' Pink’s own view that our volitional or intention-forming capacity is an "executive" rather than a purely deliberative faculty, however, has a historical origin outside Greek thought, which explains why the thick formal sense of volition inevitably involves deeper issues about the very structure of motivation. As Charles Kahn explains, the "theological concept of will," which "begins with Augustine and culminates in Aquinas and the medieval 'voluntarists'" [i.e. Scotus], models human volition on the divine will. It inherits from the biblical tradition a notion of the will as a kind of self-unifying motivation. As Kahn says, Dihle has explained the origins of this notion:

In his recent and extremely valuable Sather Lectures, Albrecht Dihle has adopted the perspective of the theological tradition. His thesis is that the concept of the will as a factor or aspect of the personality distinct from, and irreducible to, intellect and desire or reason and emotion is completely absent from the Greek tradition, but implicit from the beginning in the biblical notion of obedience to the commands of God....The appropriate human response [to divine command] is to be seen neither in terms of rational understanding, nor in terms of emotion and desire, but as a commitment of the whole person that calls out for the concept of will for its articulation.

---

136 Ibid.
138 Charles Kahn, "Discovering the Will: From Aristotle to Augustine," in *The Question of "Eclecticism:"
140 Kahn, p.236-7.
The theological tradition thus introduces a thick notion of volition as a motivation for forming intentions which is neither mere appetite nor rational intellect, but was placed by Christian apologists in what Plato had called the "middle part of the soul."

However, Kahn argues that since "Augustine’s concept of the will does not get a fully philosophical development until it is integrated within a theoretical model of the psyche, namely Aristotle’s,"\(^{141}\) we need to understand how accounts of deliberate action and choice developed from Aristotle through the Stoics to Aquinas in order to bring the theological concept to fruition. At the start of this progression,

there is no one concept that ties together the voluntary, \textit{boulēsis} or desire for the end, and \textit{prohairesis}, or deliberate desire for the means. But it is precisely the role of \textit{voluntas} in Aquinas to perform this work of conceptual unification.\(^{142}\)

Kahn describes in detail the development of this idea of \textit{voluntas} and its superiority as an analysis of will over Aristotle’s divided model. Since "Aquinas partially identifies \textit{liberum arbitrium} with \textit{voluntas} or ‘the will’ as the power to makes decisions..." He thus establishes a close connection between the will and the concept of freedom that is unparalleled is Aristotle or any Hellenistic Greek discussion of \textit{boulēsis}."\(^{143}\) Yet, while Aquinas’s \textit{voluntas} is clearly an account of second-order agency, Kahn neglects to note that, both because of Aquinas’s commitment to the Platonic principle that "we can will to do evil only if we believe that it is in some way good for us,"\(^{144}\) and because he follows the Stoic tradition in conceiving will as a rational control over appetites,\(^{145}\) Aquinas’s account loses precisely the biblical sense of volition as a \textit{distinctive} self-unifying motivation which Augustine sought to

\(^{141}\)Ibid, p.238.
\(^{142}\)Ibid, p.240.
\(^{143}\)Ibid, referring to Aquinas, \textit{Summa Theologiae}, I.83.4.
\(^{144}\)Ibid.
\(^{145}\)Ibid, pp. 244 and 247.
capture, and which Kahn detects in Epictetus as well. As he says:

Epictetus’s use of *prohairesis* serves to expand the notion of consent into the broader notion of moral character and personal ‘commitment’ as shaped by our day-to-day, moment-to-moment decisions on how to deal with our inner feelings and relationships; and this notion is presented not only as the decisive factor in practical existence but as the true self, the inner man, the ‘I’ of personal identity.\(^{146}\)

Thus in Epictetus’s notion of "moral choice" or *prohairesis* we find an anticipation of the thick motivational concept of will that existentialism requires—a concept that conceives the will as "a continual process of self-definition," by which an *individual* in radical sense irreducible to an instantiation of a rational form comes into being.

It should be noted, of course, that many philosophers reject both the idea of second-order agency and the radical sense of individuality that emerged with the thick concept of the will. In *Shame and Necessity*, Bernard Williams has argued that these ideas, which "progressivist" interpreters have found lacking in Homer and early Greek literature, "are not so much the benefits of moral maturity as the accretions of misleading philosophy."\(^{147}\) No further "unity" is needed to "have thoughts and experiences" and explain our sense of ourselves as agents than the "living person" as an animated bodily whole which we find in Homer’s characters.\(^{148}\) They make ‘decisions’ in the same sense which we ordinarily do, Williams claims: namely by acting "for reasons."\(^{149}\) As Williams argues convincingly, in most Homeric cases, intervention by the gods in human decisions simply represents the fact that "why one reason should prevail rather than another, or take over someone’s attention, can

---

\(^{146}\)Ibid, p.253.


\(^{148}\)Ibid, pp.26 and 24.

\(^{149}\)Ibid, p.32.
remain hidden.”\textsuperscript{150} Thus:

The interventions of the gods, then, operate within a system that ascribes action to human beings; and deliberations, as result of which they act; and therefore, reasons on which they act. In ascribing reasons to people, it also ascribes to them desires, beliefs, and purposes.\textsuperscript{151}

What more is needed? If there is any notion of "will" still absent here, it could only be a metaphysical illusion introduced by later philosophical corruption or ordinary language and intuitions, for "the complex net of concepts in terms of which particular actions are explained" was otherwise "the same for Homer as it is for us."\textsuperscript{152} In particular:

Homer has no word that means, simply, to decide. But he does have the notion. For he has the idea of wondering what to do, coming to a conclusion, and doing a particular thing because one has come to that conclusion, and that is what a decision is.... All that Homer seems to have left out is the idea of another mental action that is supposed necessarily to lie between coming to a conclusion and acting on it; and he did well in leaving it out, since there is no such action, and the idea of it is the invention of bad philosophy.\textsuperscript{153}

But Williams’s insistence that our philosophically unencumbered understanding of action supports this Homeric baseline is dogmatic and misrepresentative; as Pink argues instead, "everyday beliefs about decision-making," intention, and freedom point strongly in the direction of the idea that intention-formation is itself an action. The concept of the will as second-order agency is "the common-sense psychology," and it is only because "Enlightenment psychologies" that radically revise ordinary intuitions have been dominant for centuries that the thin concept of volition for which Williams finds Homer’s resources sufficient appears at all adequate.\textsuperscript{154}

\textsuperscript{150}Ibid.
\textsuperscript{151}Ibid, p.33.
\textsuperscript{152}Ibid, p.34.
\textsuperscript{153}Ibid, p.36.
\textsuperscript{154}Pink, \textit{The Psychology of Freedom}, p.28. Thus Pink dismisses Williams’s "special scorn" for "the idea that decision-making is a second-order executive agency" with the rejoinder that "What is more plausibly a (continued...)
Moreover, in theological tradition which led to the thick concept, as we have seen, will involves not merely second-order agency but also a distinctive kind of motivation for intention-formation. Williams himself concedes that "what is ordinarily called will" includes the notion of "efforts of will." He insists that Homeric characters can make such efforts, both outwardly and "within the mind," as when a man ‘dialogues with his own thumos’ in order to discover with which course of action "he is more identified." Odysseus demonstrates a capacity for "endurance" of suffering and frustration of immediate impulses to realize long-term goals, a "capacity to hold out against feeling or desire." Yet then to this extent, pace Williams, we should say that a thick notion of will as motivation by resolve —determination to persevere in maintaining intentions and plans one has formed—is present in Homer. Williams fails to appreciate this because, following Nietzsche’s reaction to Kant, he assumes that the reason progressivists find the will lacking in Homer is that his notion of action "did not revolve round a distinction between moral and nonmoral motivations." But while the thick notion of volitional motivation contrasts with the idea that beliefs and desires are sufficient to explain all motivation for our intentions, as on the Pro-Attitude model, it is not identical to moral motivation in any allegedly "abstract modern sense."

Nevertheless, the full theological notion of volitional motivation as a way unifying the self by actively forming its character, and thus of determining rather than merely discovering

---

154(...continued)
peculiarity of ‘bad philosophy’...is the attempt to conceive the freedom and agency of our decision-making in [purely] deliberative terms, as a freedom and agency of practical judgment” (pp.31-32).

155Williams, Shame and Necessity, p.36.
158Ibid, p.41.
159Ibid, p.41. The error in which Williams is caught here will become clearer in §I.10.4.
with which course of action we are "more identified," seems to be lacking not only in Homer, but in all Greek thought at least until Aristotle (e.g. *Nichomachean Ethics* III.5) and Epictetus. While philosophical attention to this real phenomenon was certainly undertaken in service of ethical themes (such as the Stoic theme of "cultivation of the self," as Foucault calls it\textsuperscript{160}) the resulting concept of volitional motivation —which was never clearly articulated— is not itself normative.\textsuperscript{161}

But in seeking a better articulation of this theological concept of will as a distinct type of motivation, we cannot be content with one traditional option: namely, volition as a merely derivative *combination* of cognition and appetite. Iris Murdoch’s account is a good example of this kind of ‘interaction model,’ or weak notion of middle-soul volition:

What moves us —our motives, our desires, our reasoning— emerges from a constantly changing complex; moral change is the change of that complex, for better or worse. Herein intellectual experiences, states of reflective viewing of the world, are constantly moving in relation to more affective or instinctive levels of thought or feeling....St. Augustine too, using a great many real-life examples, pictures will as a blend of intellect and feeling...The problem of the freedom of the will must be thought of as lying inside such a picture. Freedom (in this sense) is freedom from bad habit and bad desire...\textsuperscript{162}

Such an amalgam or mish-mash conception of will in the thick formal sense is inspired by Greek moral psychology, but remains inadequate for the theological concept of will. As Kahn and Dihle insist, for Augustine and his descendants, the will constitutes a different and *sui generis* type of motivation altogether, although it is linked both to reason and action.\textsuperscript{163}


\textsuperscript{161}Nicholas White is right that Williams’s zeal to reject all Platonic and Kantian moral idealism leads him to distort the archaic Greek world-picture in a Nietzschean direction: see White, Review of *Shame and Necessity*, *Journal of Philosophy*, Vol. 91.11 (November 1994): 619-622.

\textsuperscript{162}Murdoch, p.300.

\textsuperscript{163}Ultimately, this idea derives from the doctrine of the *imago dei*, which implies that human beings have a capacity for motivation *like* that which is expressed in God’s act of creation *ex nihilo*—an act which is not rationally required or desired, as if God would be incomplete without the world, but instead an *overflowing* of (continued...
In this tradition, freedom is the liberty of this self-motivating movement, rather than the liberty of mere decisions about this or that outward act, but then such liberty cannot be reduced to the habit of rightly guiding appetite by reason either. On the classical picture such a phenomenon could only be located in the ‘middle-soul.’ But if we are to avoid the confusions which this placement can inspire (as Murdoch’s case shows) we must more carefully consider what other phenomena the ‘middle-soul’ might include, and how these can be distinguished from the will in the thick sense.

I.4. Sentiment, Will, and the Ambiguous ‘Middle Part of the Soul’

C.S. Lewis sheds an interesting light on this matters in his remarkable little book, *The Abolition of Man*. This work starts as a response to what Lewis sees as an emotivist trend in humanities textbooks for schoolchildren: through their interpretation of examples, they teach "firstly, that all sentences containing a predicate of value are statements about the emotional state of the speaker, and secondly, that all such statements are unimportant."164 The textbook authors "see the world around them swayed by emotional propaganda” and seek to protect students against it by entirely numbing their capacity for what moral psychologists today call ‘evaluative attitudes’: they fail to see that "the right defence against false sentiments is to inculcate just sentiments."165 Lewis’s fear is that children will learn from these books that communicated emotions are purely ‘subjective,’ reflecting only the brute preferences of their bearers; lacking objective communicable content, they involve no judgment that could be in

163(...continued)
His perfection.


any sense right or wrong.\textsuperscript{166} If all figurative language is denied anything but emotivist significance, then the responses it arouses cannot be more or less \textit{appropriate} in any objectively meant sense, and the child is robbed of "experiences which thinkers of more authority than [the authors] have held to be generous, fruitful, and humane."\textsuperscript{167} As Lewis says,

Until quite modern times all teachers and even all men believed the universe to be such that certain emotional reactions on our part could be either congruous or incongruous to it—believed, in fact, that objects did not merely receive, but could \textit{merit}, our approval or disapproval, our reverence, or our contempt....The man who called the cataract sublime was not intending simply to describe his own emotions about it: he was also claiming that the object was one that \textit{merited} those emotions.\textsuperscript{168}

Because there is "objective value" to which our attitudes respond, "emotional states can be in harmony with reason." Yet such evaluative attitudes, as they are now called, are not ‘speculative’ cognitions or thoughts with straightforward propositional content:

\begin{quote}
No emotion is, in itself, a judgment: in that sense all emotions and sentiments are alogical. But they can be reasonable or unreasonable as they conform to Reason or fail to conform. The heart never takes the place of the head: but it can and should obey it.\textsuperscript{169}
\end{quote}

As we see from this familiar metaphor of the ‘heart,’ which derives from the \textit{Republic} 441b where Plato quotes Homer to show that the \textit{thumos} or ‘spirit’ is distinct from the reasoning part of the soul,\textsuperscript{170} by "sentiment" Lewis means the ‘middle part of the soul’ in the Platonic

\textsuperscript{166}Ibid, p.19: "What he will learn quickly enough, perhaps indelibly, is the belief that all emotions aroused by local association are in themselves contrary to reason and contemptible."

\textsuperscript{167}Ibid, p.20.

\textsuperscript{168}Ibid, p.25.

\textsuperscript{169}Ibid, p.30. This view would have to be reconsidered in an analysis of the relation between emotional and volitional states.

\textsuperscript{170}Plato, \textit{Republic}, 441b-c: "And to these instances we may add the testimony of Homer quoted above: ‘He smote his breast and chided his heart’ [\textit{Odyssey}, 20.17]. For there Homer has clearly represented that in us which has reflected on the better and the worse as rebuking that which feels unreasoning anger, as if it were a distinct and different thing." It is notably that Plato’s bodily allegory for the tripartite soul, which links reason with the head, spirit with the heart, and appetite with the hands, has retained a continuing influence down

(continued...)
tripartite psychology. It is this middle part, the ‘heart,’ which Lewis fears is excluded by the Humean ontology our age has inherited: if all that exists are ‘valueless’ facts apprehendable only in thought and ‘factless’ values that correspond to unarticulable subjective preferences, then the ‘sentiments’ or evaluative attitudes have no objects.\textsuperscript{171}

In Lewis’s view, then, the middle part of the soul consists in affects or feelings that, unlike certain appetites, are not instinctively determined by our biological nature, but are malleable and can be trained to respond to value-insights.\textsuperscript{172} And in both Plato and Aristotle’s moral psychologies, the right education of these dispositions is essential to virtue:

Without the aid of trained emotions the intellect is powerless against the animal organism. I had sooner play cards against a man who was quite skeptical about ethics, but bred to believe that ‘a gentleman does not cheat,’ than against an irreproachable moral philosopher who had been brought up among sharpers....We were told it all long ago by Plato. As the king governs his executive, so Reason must rule the mere appetites by means of the ‘spirited element.’ The head rules the belly through the chest—the seat, as Alanus tells us, of Magnanimity, of emotions organized by trained habit into stable sentiments. The Chest—Magnanimity—Sentiment—these are the indispensable liaison officers between cerebral man and visceral man. \textit{It may even be said that it is by this middle element that man is man.}\textsuperscript{173}

From a scholarly perspective, this account (written for the layperson) is undoubtably simplistic. Though it employs tripartite imagery, it does not consider the highly controversial question of whether there is a single unified sense of \textit{thumos} in Plato’s

\textsuperscript{170}(...continued) through the centuries. Even in the end of Fritz Lang’s famous expressionist film, \textit{Metropolis}, for example, the last scene ends with the caption: "There can be no understanding between the hands and the brain unless the heart acts as mediator."

\textsuperscript{171}Lewis, p.30-31: "On this view, the world of facts, without one trace of value, and the world of feelings without one trace of truth or falsehood, justice or injustice, confront one another and no \textit{rapprochement} is possible."

\textsuperscript{172}In this sense the might be compared to Hume’s "artificial sentiments," except that Lewis does not conceive the moral distinctions to which they respond as largely conventional or human artifacts.

\textsuperscript{173}Lewis, p.33-34, my italics. Hence, using the metaphor of the heart for trainable sentiment, Lewis says that the education which numbs sentiment produces "Men without Chests" (p.34), who lacks precisely what is distinctive of \textit{humanity}. 

Page -45-
writings, nor how Aristotle developed and altered this idea of a middle-soul in his concept of the ‘intellectual appetite.’ Lewis also neglects the complex way in which these notions served as the historical antecedents of the notion of ‘will’ in later Christian philosophy of late antiquity and the middle ages. As a result, Lewis does not ask whether trainable ‘evaluative attitudes’ give us a sufficient account of volition, or whether the classical notion of a ‘tripartite’ division can be made philosophically precise and stand up to scrutiny—a question that requires analysis of the concept of ‘emotions,’ how they differ from (other?) desires and appetites as well as states of cognition, and whether volition should be considered some subclass of them or something else. But although it leaves all of this large agenda unaddressed, Lewis’s suggestion that it is the middle-soul which is uniquely human contributes something valuable. If we read his proposal to mean that ‘it is by the will that human persons are persons,’ then it expresses the intuition underlying this whole dissertation.

Although Lewis frames his discussion of the middle-soul in terms of evaluative attitudes, it is clear that "sentiment" in his sense has a motivational aspect. These sentiments not only have objects and involve practical judgments, but motivate us to frame intentions and to act on them, which is why ‘right sentiment’ (in accord with ethical principle) is the basis of virtues. Without them, "no justification of virtue will enable a man to be virtuous." This link between evaluate sentiment and virtue has suggested to many thinkers that the ‘will’ is just the function whereby the middle-soul conforms our motivation and decisions to the results of practical deliberation. The will is thus determined entirely by the intellect.

This first suggestion, which derives from Aristotle’s account of choice in the

\[\text{\textsuperscript{174}}\text{Lewis, p.33.}\]
Nicomachean Ethics III.3 as "deliberate desire of things in our power,"\textsuperscript{175} offers a proleptic characterization of ‘will’ that remains neutral on the issue of ‘freedom’ and merely circumscribes the domain in which we are to look for a more adequate conception.

Following this first functional outline of the will, Lewis’s short exposition would leave us with the image that volitional states are a kind of combination of reason and raw appetite, in which reason becomes practical and appetite becomes trainable feeling by a sort of ‘blending’ (as we saw in Murdoch’s conception). I will call this the composite approach to the will in the thick sense. It is a general ‘approach’ that includes or covers several different specific conceptions of volition. Conceptions taking the composite approach are arguably more indebted to what most Aristotle scholars describe as his bipartite moral psychology, in which there are only two fundamental categories in moral psychology —intellectual and appetitive— and all other types of mental phenomena are analyzed as subsets of or relations between these categories. The composite approach is not genuinely tripartite, since it does not recognize a middle-soul that is of equal primordiality with reason and appetites or inclinations, but holds that phenomena ascribed to the middle faculty can be ‘decomposed’ into other constituents.

Writers the philosophy of religion indebted to the Thomistic tradition often hold bipartite conceptions of this sort. For example, in a recent essay, Hugh McCann defends the view that ideally, the principle of sufficient reason will cover human choices and decisions.\textsuperscript{176} The common view advanced by McCann, namely that it is philosophically more satisfying if there is (at least in principle) a "full explanation" for our choices —and that libertarianism is

\textsuperscript{175} Aristotle, *Nicomachean Ethics*, 1113a11.

unsatisfying if it entails that there can be no sufficient explanation for free decisions\(^\text{177}\)— is a reflection of the analysis of choice in terms of practical reasoning which McCann inherits from Aristotle and Aquinas. Forming an intention to act a certain way—e.g. to go skiing in Colorado—is described by McCann as a "mental act" which follows from a practical syllogism beginning with the premise "Would that I go skiing."\(^\text{178}\) As McCann says, "the major premise is optative; it expresses, not a belief on my part that a skiing trip would be desirable, but rather the actual desire with which I experience the prospect of such a trip."\(^\text{179}\)

But is simply desiring or having a preference in this sense the same as actually willing to go skiing, from which my will to go to Colorado is derived? McCann quickly adds that the optative state expressed by the major premise is still a state of thinking:

For to think anything in the optative mode is to perceive it as in some way good, and we decide in light of our optations. The upshot of this is that we form intentions only in light of a perceived good.\(^\text{180}\)

This is simply the ‘Platonic Postulate’ again, or the thesis that all motivation is driven by our apprehension of the good, and so the will characteristic of persons is reducible to nous. Note that by beginning with this presupposition, McCann entirely bars from consideration the claim that really makes the Kantian form of libertarianism original.\(^\text{181}\) Kant’s "Incorporation thesis"\(^\text{182}\) assures us that an optation which constitutes a will for some end involves something more than the purely mental state of finding that end attractive or experiencing an

\(^{177}\) I will challenge this thesis in the projected final chapter on free will.


\(^{179}\) Ibid, mss. p.2.

\(^{180}\) Ibid, mss. p.3.

\(^{181}\) And since it is Kant’s version of libertarianism which has been attacked as violating the principle of sufficient reason, this is the version which is most relevant for the question McCann is considering.

\(^{182}\) I am using Henry Allison’s name for the thesis Kant gives in the "Observation" section at the beginning of Book One of the Religion within the Limits of Reason Alone (p.19 in the Harper Torchbooks edition translated by Theodore Greene and Hoyt Hudson).
impulse to realize it: willing that end requires a movement of negative liberty which makes the desire into one’s maxim. McCann, by contrast, reduces decisions to conclusions from optations, which reduce in turn to desires resulting from mental acts of apprehending the value of certain ends.

This kind of motivational account of volition contrasts with another which —often through linking it to more radical kind of liberty, as the Kantian example suggests— the middle-soul and its ‘freedom’ are analyzed as irreducible to any relation between reason and appetite, such as dispositions of the latter being habituated to respond to the former. If it is still understood as involving ‘desire,’ this is distinguished in the middle-soul from ‘lower’ forms of impulse or attraction.

The historical differences made by these two distinct approaches to the will as a thick motivational phenomenon are perhaps often more easily seen in the treatment of another directly related question: namely, the structure of ethical virtue, which Aristotle analyzed roughly in terms of dispositions of his (supposedly composite)\textsuperscript{183} intellectual appetite, but which later scholastics such as Duns Scotus thought of as qualities of the will alone, as Bonnie Kent has shown in a recent study.\textsuperscript{184} In discussing virtue, Lewis naturally sticks with his language of trained affects:

St. Augustine defines virtue as ordo amoris, the ordinate condition of the affections in which every object is accorded that kind and degree of love which is appropriate to it. Aristotle says that the aim of education is to make the pupil like and dislike what he ought. When the age for reflective thought comes, the pupil who has been thus trained in ‘ordinate affections’ or ‘just sentiments’ will easily find the first

\begin{footnotesize}
\textsuperscript{183}I think there are serious reasons to dispute whether Aristotle did in fact ultimately propound a bipartite moral psychology in which intellectual appetite is decomposable, but this issue requires much more detailed treatment than I can give it here, and does not change the fact that Aristotle’s moral psychology was and still is generally received as bipartite.

\textsuperscript{184}See Bonnie Kent, Virtues of the Will: The Transformation of Ethics in the Late Thirteenth Century (Catholic University Press, 1995), reviewed by Philip Quinn, Journal of Philosophy, 93.12 (December, 1996): 628-631. As Quinn puts it, “for Scotus, strictly speaking, all virtues belong to the will” (p.628).
\end{footnotesize}
principles of Ethics.\textsuperscript{185}

But it is not obvious that virtues should be analyzed as desires or sentiments built by "trained habit," in Lewis’s term, if they are instead dispositions \textit{of choice}, or qualities of the will. This issue was central to medieval debates about virtue. For example, as Philip Quinn summarizes:

Kent charts the development from Aquinas, who located courage and temperance, which order the passions, in the sense appetite, to Scotus, who located all the virtues in the will....Appealing to Aristotle’s view of virtue as a habit concerning choice, Richard of Middleton contended that virtues belong to the will because choice is an act of the will. Conceding that the sense appetite develops habits of obeying the will, he described such habits as mere likenesses of virtue. Olivi argued that only the will has virtues \textit{simpliciter}, habits in other powers being virtues only in connection with the related virtues of the will.\textsuperscript{186}

This debate raises the difficult issue of what different sorts of ‘disposition’ there might be, and whether, for instance, we should clearly distinguish ‘habits’ as dispositions to \textit{behave} in certain characteristic ways, or even as propensities to act on certain sorts of intentions, from dispositions of \textit{volition}, i.e. dispositions to be motivated to decision in certain ways, or to form certain kinds of intentions rather than others. ‘Volitional dispositions’ in this sense are importantly different from patterns of accustomed first-order actions, or behavioral traits.\textsuperscript{187}

These cursory reflections suggest the possibility that evaluative attitudes and volitions are \textit{distinct}, even if each is (in part) ‘dispositional.’\textsuperscript{188} But if this is so, then our initial reference to ‘the’ middle part of the soul has to be made less monolithic. There are in all probability several \textit{different kinds} of psychological phenomena that we should distinguish both from

\textsuperscript{185}Lewis, p.26.

\textsuperscript{186}Quinn, Review of Bonnie Kent, \textit{Virtues of the Will}, op. cit., pp. 630-1.

\textsuperscript{187}As Stanley Hauerwas points out, for Aristotle, virtue is "a habit that is formed by activity" involving choice, not by mere rote repetition, and thus requires "an established \textit{tendency to act} from deliberate decision, a mode of choice" —see Hauerwas, \textit{Character and the Christian Life}, p.71. Disposition in this sense has to do with our motivation to act in virtuous or vicious ways, rather than simply with stability in our pattern of outward behavior.

\textsuperscript{188}Though we have yet to discern what dispositionality itself might consist in.
lower inclinations or appetites at one end of the spectrum, and from pure cognitions or speculative judgments with objective propositional content at the other end. For example, recently a whole literature has arisen on the ambiguous topic of ‘emotions’ which, at least since Sartre’s famous treatment,\(^{189}\) has raised them from the status of mere chaotic feelings or even epiphenomenal consciousness of physiological disturbances\(^{190}\) to states with purposive functions and unique intentional meanings, or even rational judgments, on some accounts. But in this literature, emotions are often not clearly distinguished from other states, such as desire and cognition, let alone from the two other potential residents we have identified in the ‘middle soul’: volitions and evaluative attitudes. It is likely that there are relevant distinctions to be drawn here. Even though Lewis described the "sentiments" or evaluative attitudes as "trained emotions," and gives them a cognitive significance denied by emotivism, he denies that they are simply equivalent even to practical let alone speculative judgments.

Romantic thinkers influenced by German idealism had a tendency to celebrate the middle-soul while obscuring these important distinctions within it. Coleridge, for instance, tries to link "thinking" and "feeling" in a general or undifferentiated way, "expanding the concept of Reason to include the un-Kantian elements of emotion and need."\(^{191}\) As Wayne Anderson explains, for Coleridge, although Reason divides into the speculative and practical, only the latter, which includes imagination and understanding of concrete sensory experience, is fully human:


\(^{190}\)This is William James’s famous "peripheric theory" of emotions, according to which, for example, someone is sad because he weeps, rather than the reverse.

Though he divides Reason into two parts, he generally means it to denote the province of practical reason and thus of concrete human choices and needs. For practical reason contains all the faculties of men, ‘comprehending the will, the conscience, the moral being with its inseparable interests and affections.’...The objective rationalists maintain that only the object —devoid of the beauty and value we bring to it— is real. The mind for them is rational intellect alone, evacuated of desire. Coleridge maintains, in contrast, that genuine Reason ‘never acts by itself. but must clothe itself in the substance of individual Understanding and specific Inclination.’...Reason is consequently united with understanding; it is not exclusive of rationality, but in excess of it. True knowledge is informed by all the faculties. Reason synthesizes will, emotion, conscience and intellect, representing the claims of the whole man.192

Anderson points out that this embracing concept of reason influenced the Cambridge Platonists,193 who opposed "rationalistic reductions of the mind" and held that faith was both rational and "derived fundamentally from the will and heart."194 In this they anticipated Coleridge’s declared view that "'Faith subsists in the synthesis of the reason and the individual will'"195—a view which certainly anticipates later existentialists.

In the next chapter, however, I will try to show that it is possible to avoid running together feeling, desire, emotion, will, and practical evaluation, while still remaining true to Lewis’s and Coleridge’s intuition that the essence of personhood resides in the capacities of the 'middle soul.' The philosophical potential of this intuition can be realized only through opening up the image of "sentiment" or "heart" to the complexities that lie within it.

I.5. Towards a Preliminary Moral-Psychological Taxonomy


193I think it apparent that it also influenced C.S. Lewis, whose ideas on sentiment owe much to Coleridge and his fellow romantics.


In contemporary philosophical and psychological literature as well, we find widely
variant uses of terms such as ‘impulse,’ ‘inclination,’ ‘appetite,’ ‘affect,’ ‘desire,’ ‘feeling,’
‘emotion,’ ‘disposition,’ ‘mood,’ ‘belief,’ ‘volition,’ and of course ‘reason’ and ‘decision;’
as a result, confusion abounds in moral psychology. We must cut a path at least some way
through this thicket before the question of whether there is a distinctive sense of will or
volition can even be framed intelligibly, and its relative absence from contemporary moral
psychology clarified. In the next chapter, I will pursue the intuition that we can approach a
proper interpretation of the will indirectly, through showing how volition can be
distinguished from certain kinds of ‘desire’ with cognitive content, from the practical
deliberation that accompanies them, and finally from emotions, all of which can claim a kind
of middle-role in our psychologies between the operations of ‘appetite’ and ‘intellect,’ as the
poles were called in classical terminology. This in turn requires indicating at least in broad
outline how these ‘middle-soul’ psychological kinds differ from other related phenomena,
such as feeling, appetite or lower sorts of desire, and forms of judgment— in other words, it
requires sketching a provisional hierarchy of different kinds of motivational states.

Two objections. First, it can be protested that we should not just assume that any of these
psychological concepts are mutually exclusive in meaning, or that the range of one does not
intersect with or form a subset of the ranges of some of the others. Second, someone may
object that the potential overlaps between these concepts are at least in part ‘merely
semantic,’ since they will vary accordingly to whether we employ narrower or wider senses

---

196 The format of the categorial problem in moral psychology appears to be typical of any important
conceptual analysis. We are told that we can either start by examining ‘paradigm cases’ in the extension of
each concept, and revise our concepts accordingly, or start by spelling out the concepts themselves, and
reclassify our instances accordingly. The hermeneutic methodology which I embrace, however, does not accept
this rigid dichotomy: it holds instead that paradigm cases were always taken as such implicitly in the light of
some conceptual interpretation, and that the concepts themselves were always at least implicitly understood in
terms of their typical applications, so that both must be adjusted together in the process of distinguishing
concepts.
of ‘desire,’ narrower or wider senses of ‘emotion,’ narrower or wider senses of ‘feeling,’ and so on, since all these terms are multivalent and have different scopes in different usages. My response is that for my present purposes, it will be most useful to try to articulate narrower senses of these concepts. Thus to the (limited) extent that the question of overlap is merely semantic, our heuristic should be (as far as is reasonably allowed by the relevant phenomena themselves) to pick senses of desire, emotion, volition, and so on that are as restrictive as necessary to avoid or minimize overlaps, so that each term acquires its most distinctive permissible sense. This will help in isolating the special motivational role of volition.

Following both this heuristic and the general hermeneutic paradigm of concept-analysis, I will begin in the next chapter by distinguishing several types of ‘desire,’ and then argue that there is a kind of human motivation which is not desiderative in any of these senses. In light of these distinctions, we can give specific content to a ‘middle-soul’ or thick and motivational notion of will.

A final methodological remark. Since most of the ‘folk psychological’ terms of moral psychology have multiple and conflicting senses, as we have already seen, it is one thing to consider where natural divisions in the phenomena lie, and another to determine which names to use for which. If everyone could agree on the former, then the later would only be a matter of convention. But because the connotations associated with the terms help define the relevant concepts and in turn their paradigm instances, a pure phenomenology ‘carving at the joints’ and assigning conceptual terms later cannot be attained: our best option is a ‘hermeneutic’ phenomenology that works from both ends at once. Nevertheless, for the sake of clarity, in addition to traditional terms I will also use a set of my own labels for different types of psychological states (S1, S2, S3... etc.), which will allow me, in certain contexts, to
avoid relying solely on the ambiguous connotations of the familiar terms. The series of states I will describe is meant to constitute a hierarchy in the sense that animals which are capable of any given state in the list generally also possess states lower down the list, though not because the states of a higher kind always include or involve states of the lower kind.

My strategy in Chapter II will be to lay out a *partial and provisional* classification of the relevant types of psychological states, culminating with a distinctive kind of volitional motivation. Then in Chapter III, the resulting analysis of volition will be refined through discussions of particular themes that illustrate volitional motivation in Frankfurt, Levinas, Scotus, and Arendt.

---

197 By this, I do not mean to be committing myself to any sort of psychological atomism. There may be complex states that incorporate others and yet rise beyond their composite properties in new syntheses, in which they still count as new primitive kinds. Some of those kinds in the middle of the hierarchy may well turn out to be of this nature.

198 As this series of labels suggests, the underlying assumption of my approach is that the phenomenology and psychology of inner life reveal a series of different kinds of states from, speaking very loosely, completely irrational unconscious instinctive behavior to maximally abstract, detached intellectual speculation. Thus, however its components are divided, the field studied by moral psychology forms some sort of a *hierarchy*, rather than a mere collection of different states that can be arranged in any order. Still, this does not mean that one is 'higher' than another in an ordinal or cumulative sense, such that, for instance, S3 states include S2 states and something more. Rather, in keeping with the heuristic stated above, the hope is to identify the sparse set of 'primitive' or non-overlapping kinds or levels of psychological state, out of which more complex states are composed.

199 The latter is the sense in which Aristotle meant that the particular types of 'soul' subsumed under the common term "constitute a series, each successive term of which potentially contains its predecessor..." See Aristotle, *De Anima*, tr. J. A. Smith, in *The Basic Works of Aristotle*, ed. Richard McKeon (Random House, 1941), II.2, p.560, 414b30-31.
CHAPTER II

Desire and Will:

The Structure and Existential Role of Projective Motivation
Contents of Chapter Two

II.1. Introduction

II.2. Three Kinds of Desire

II.3. A Preliminary Interpretation of Desire as ‘Lack’
   - II.3.1. Desires vs. Evaluative Attitudes
   - II.3.2. Plato’s Analysis of Desire
   - II.3.3. Some Further Points From Aristotle

II.4. Nine Types of Second-Order Desire

II.5. Will as Conative Projection
   - II.5.1. Motivation vs. Intention
   - II.5.2. Projection as a Different Kind of Motivation
   - II.5.3. Virtue, Goods Internal to Practices, and Projection
     - Aristotle on Virtue and the Paradox of Eudaimonism
     - Friendship and Virtue
     - MacIntyre and the Entry into Practices

II.6. The Existential Significance of Projection
   - II.6.1. Frankfurt on Meaning and Final Ends
   - II.6.2. The ‘Good’ of Projection vs the Grounds of Projection

II.7. Kant and Projective Motivation
   - II.7.1. The Limitations of Deontic Projection
   - II.7.2. Non-Deontic Grounds for Projection
   - II.7.3. Separating Projection from Accompanying Desires

II.8. Projection and Libertarian Freedom: A Concluding Note
II.1 Introduction

In the previous chapter, I introduced the distinction between thick and thin concepts of volition, and argued that the history of ideas on volition was and is still dominated by thin conceptions that pay little attention to the motivational structure of willing in paradigmatic experiences such as inward resolve, determination to pursue some course, or personal commitment to a cause. We also saw that there is an alternative tradition, stemming ultimately from Jewish and Christian sources, of substantive accounts of willing that pay attention to its motivational qualities, and that this tradition broadly speaking analyzed volitions as "middle-soul" states, with both cognitive and appetitive aspects. But as a result, this tradition implicitly accepted the implication of Aristotelian bipartism that the will is a composite entirely reducible to non-volitional components; it also failed to draw any clear distinctions between other states that could be regarded as belonging to the "middle-soul" (spirit or heart), such as emotions, evaluative attitudes, and desires informed by or habituated to conform to practical reason. Indeed, in the Aristotelian tradition that dominated Christian philosophy prior to Kant, it was assumed that "will" meant precisely desires of this practically rational kind.

This chapter will provide a basis for a different substantive yet anti-Aristotelian account of the will as a thick motivational phenomenon. In particular, after discussing Plato’s and Aristotle’s analyses of desire and distinguishing several types of desire that are not always sufficiently distinguished in modern neo-Humean accounts, I will argue that states which are properly considered volitional (in the thick sense of this term) have a non-teleological structure that contrasts with every type of desire, including the practically rational appetitive states used to explain volition in Aristotelian accounts. This structure, which I call projective motivation, is implicit in Kant’s moral psychology, but Kant does not recognize it
as a general structure with a set of possibilities wider than deontic moral motivation (which is only one kind of projection). It is also implicit in some of Harry Frankfurt’s work on motivation and existential significance, and (as we will see in the next chapter) in his concept of caring.

II.2. Three Kinds of Desire

As I already noted, the base level (S0) of a modern hierarchical moral psychology should be unconscious behavior that is driven purely by instinct. While older hierarchies starting with Aristotle’s distinguished the "vegetative" (or nutritive) and "locomotive" souls, this was only because they assumed that no animal could exhibit motive behavior without sensation, and that sensation is always conscious. Hence Aristotle thought that local movement at least required the sensation of touch, and thus an awareness of pleasure and pain that could motivate a desire (the lowest species of appetite), since some kind of (conscious) appetite is needed to cause bodily movement. The "nutritive faculty" cannot be the source of local movement, because movement is always for an end and is accompanied by either imagination or appetite; for no animal moves except by compulsion unless it has an impulse towards or away from an object.

As Professor Lear summarizes Aristotle’s explanation in the De Anima:

Human action is a species of animal movement. All animal movement, Aristotle argues, must flow from desire. Lesser animals have basic appetites as well as

200 See Aristotle, De Anima, II.3, p.560, 415a6-7: "Again, among living things that possess sense, some have the power of locomotion."

201 In general, it seems that the problem of consciousness, as opposed to powers of soul and the epistemic features of sense and mind (nous), simply does not even occur to Plato, Aristotle, and their predecessors as a separate problem. Something like the notion of subjectivity is included in the Eleatic distinction between appearances and reality, and it motivates later skepticism, but consciousness as a phenomenon does not seem to get distinct treatment until Augustine.

202 Aristotle, De Anima, III.9, 432b 15-20. Also see III.10, 433a, and III.11, 434a 1-2.
sense perception, and imagination based on their sensory awareness. But their movements would be incomprehensible on the basis of sensation and imagination alone... 203

[Note that Lear is using "desire" for appetite in general and "basic appetites" for what our translation renders as "desires"]. 204 Aristotle did not perceive that some animals move without any kind of appetite at all, because he could not know about cases of so-called tropistic behavior that are controlled by a simple mechanism or program that goes awry and produces aberrant or repeated behavior if its routine is disturbed. 205 Although it need not be, tropistic behavior is usually interpreted as entirely unconscious in animals —such as insects— in which it predominates. 206 This makes it as much an S0 state as the "hectic activity" and adjustments constantly going on in unconscious, automatic systems such as digestion. 207 Aristotle could of course say that such behavior is compelled, rather than movement in his sense, but by compulsion he seems to have in mind external force acting on the body (as in his definition of the involuntary). 208


204But see Lear, Aristotle, p.142, note 110, where he argues that the Oxford translation is misleading, since it uses ‘appetite’ both for lower urges (epithumia) "and for the word that is standardly translated as ‘desire’ (orexis)" which is more general. The Smith translation which I am using appears to employ "appetite" throughout for the general concept (orexis) and "desire" for lower appetites, which is the inverse of Lear’s translation. In order to avoid having to transpose "appetite" and "desire" in every quotation from Aristotle to be consistent with Lear, I have chosen to stick with the Smith translation as is, and note Lear’s different usage where appropriate.


206Although the wasp’s behavior is triggered by various sensations, it is not unreasonable to assume that there is no phenomenal experience of these sensations, i.e. that they are mere reactions without there being anything that ‘it is like’ to be the wasp.

207See the informative discussion in Dr. Jonathan Miller, The Body Question (Vintage Books, 1978), p.39: "Our nervous system is designed to emphasize what’s going on in the outside world... The inside of the body —blood vessels, heart, intestine, lungs, and bladder, is literally studded with instruments capable of registering changes in pressure, temperature and chemical composition. But none of these meters has any dials: they are not meant to be read by human consciousness, but are linked up with reflex systems which obey automatically."

At the next level (S1), we have a whole range of experiences that are both instinctive and linked to physiological needs, but are conscious. We might designate these phenomena as ‘brute urges’ or impulses that are either intrinsic to the proper biological functioning of the body, such as hunger, thirst, sexual attraction, and perhaps certain kinds of disgust or aversions (e.g. to what smells or tastes foul), or aberrant or unusual versions of such inclinations, or instinctively driven conscious feelings, such as panic in a flight response. In themselves, these urges or inclinations seem to function as dispositions to behave in characteristic ways, but as dispositions, they remain pure from cognitive content. Like basic sense-qualia (such as colors and sounds), they are consciously experiencable without any connected linguistic thought or propositional attitudes. On this hypothesis, there is a discernible level of ‘feelings’ that are purely qualitative states, although they are closely connected with physiological reactions. For example, the newborn infant’s first instinctive urge for milk is conscious but probably without any more articulation than the feeling of an undifferentiated need—though it soon learns to associated this with whatever sources of milk are provided. However, what distinguishes S1 impulses is not simply the lack of linguistic articulation but their failure to have any intentional content at all: the infant’s urge is not towards anything represented as potentially satisfying it, but a pure feeling of need that is instinctively directed to the right source of satisfaction.

I will use ‘impulses’ as a term of art for such S1 states. If we want to call any S1 states "appetites" or "desires," then certainly appetites in this impulsive or purely noncognitive sense can be distinguished from another kind of state that combines the qualitative feel of the instinctive urges with a real ‘intending’ of particular objects under a certain aspect, e.g. as likely to satisfy the urge. In these cases, the intentional content directs the urge towards

---

208 (...continued)
"Those things, then, are thought involuntary, which take place under compulsion or owing to ignorance..."
something interpreted as its end: e.g. a desire for a beer, rather than raw thirst. Because of they have mental content, these (S2) states can also become the basis for an intention which guides bodily movements, such that the intention and movements together constitute an ‘action’ in the most basic sense (lifting a beer out of the refrigerator). Furthermore, by connection with instrumental reasoning, such an intention can anchor a plan of action, or an "intention" that guides a whole series of bodily movements and interactions with others in the environment that are calculated to reach the (S2)-‘desired’ end (e.g. driving to the store to buy some beer).

We can also think of such S2-desires as preferences, meaning that they express the agent’s comparative ranking of one object over others that might be candidates for satisfying the desire. Since some S2 inclinations are the result of S1 urges being directed to specific intentional objects —as in thirst for a beer— we can think of the S1 state as an unsaturated urge towards a range of possible satisfying objects, and the corresponding S2 state as a ranking of the objects within that range. But S2 states can also develop in relative independence of any underlying instinctual drive, the type of content intended in these states determines the relevant range of possible objects to be ranked. For example, I may have a very crude musical inclination for classical over pop, or a very refined preference for Mozart over Beethoven over Brahms..., etc. In every case, however, preferences amount only to what Charles Taylor has aptly termed "simple weighing." The only ‘evaluation’ of the object of desire involved in a preference is its subjective placement on a scale relative to other potential objects of inclination. As a result, the multiple inclinations which a subject has cannot conflict essentially because their contents involve no contrastive evaluations. An

S2-desire to swim and an S2-desire to eat cake can conflict only for reasons such as that eating cake gives me a cramp that prevents me from swimming: "Not being contrastively described, these two desired consummations are incompatible, where they are, only contingently and circumstantially."  

As Taylor notes, however, S2-desires have sufficient content to ground what he calls "weak evaluation" aimed at weighing desires to discover their contingent incompatibilities and how to maximize one’s overall preference-satisfaction. And where there is no conflict of preferences, an S2-inclination can sometimes be sufficient for an instrumental plan to guide a whole course of action. Consider an example in which an S2 state (or ‘preferential inclination,’ as I will call them) results from raw sexual impulse being directed towards a certain potential mate named Cheryl. Matthew conceives a desire ‘to go out with Cheryl and get to know her better.’ Of course, reasons and motivations of quite a different nature certainly might enter into the formation of such an intention, but at this point we are hypothetically imagining a case in which all the agent’s acts towards Cheryl are done, so to speak, with ‘only one thing in mind,’ and are motivated solely by their instrumental relation to the biological impulse for a sexual union. We still recognize that in such a case, the agent has a plan of action and particular inclinations or desires to do things that are likely to advance this plan (e.g. the desire to ‘call Cheryl tonight’). These cases constitute ‘desires’ in an instrumental extension of the S2 sense, and so remain on the same level.  

Inclinations or desires in the S2 sense are thus ‘original’ when they stand at the top of the instrumental chain motivating other more particular desires. Of course, this does not require that the agent has such a plan explicitly articulated and mapped out for himself: the relation of one S2 desire to desires for particular things that promote the satisfaction of the first

---

However, I do not mean to import into the definition of S2 desires anything like the formal apparatus of restrictions on preferences which rational decision theorists have evolved. But in every case, the S2 desires or inclinations are distinguished by involving mental contents that are purely subjective in the following important sense: they involve no evaluations that make a validity claim on others looking for their agreement, since they do not expect any kind of interpersonal ratification, or even (in more unusual cases) intelligibility to third-parties. My desire to listen to Mozart’s Eine Kleine Nachtmusik does not imply that other should enjoy hearing it as well: if my desire is motivated by reasons about the music’s qualities which I think others should recognize, then it is not an preferential inclination. In Matthew’s case (assuming that he has moral or at least prudential reasons to ignore other strategies), he needs Cheryl to understand his desire to go out with her so that she can agree to the date, but his communication of this to her is purely strategic: he does not require her to think that she has good reasons to go out with him, or that he is in some objective sense worthy, but only that she share the same preference for a date. Thus, in most cases at least, the content of S2-desires is communicable, but if the S2-desire itself motivates the communication, it can only be strategically communicated. In these respects, all S2 states function like personal preferences in rational decision theory: one person’s preference for chocolate ice cream involves no claim whatsoever about what flavor of ice cream another person should prefer. My telling the soda-shop waiter that I would like chocolate is (considered by itself) just strategic action —though I might do it in a polite way, for other reasons.

Even though they are routinely communicated, then, S2 ‘preferential desires’ (as I will call them) involve no ‘validity-claim’ or implicit reference to ideal standards of warrant or normative burden to provide reasons for the claim to others in a critical exchange. There is

\footnote{However, I do not mean to import into the definition of S2 desires anything like the formal apparatus of restrictions on preferences which rational decision theorists have evolved.}
not even always a requirement that others be able to understand what the preference means to its owner, unless this is required (as is often the case) by the derivative instrumental need they have for others to help them in satisfying it. For example, I need the chef to understand my preference for ‘medium-well and-peppery’ so that he knows how to cook my steak the ‘way I like it.’ But if the chef is precocious enough to ask me why I want my steak this way, when everyone of cultured taste knows that it is best medium-rare, since this allows the full flavor to emerge, etc. etc., there is simply no possible answer I could give. My preference, though it is cognitively specific rather than impulsive, is not motivated by any (conscious) reasons, and so it is not on the same level as the type of aesthetic evaluation with which the chef is challenging it. A being who was only capable of S2-states would be as deaf to the chef’s appeal as my cat is to my frustrated plea that his pill is good for him in an objective biological sense.

Of course, since we are not like the cat in this respect, in practice S2-inclinations may almost always be closely connected with other states that are not subjective in the S2 sense: e.g. Cheryl’s belief that Matthew is ‘cute,’ which motivates certain reactive attitudes in her, such as disagreement, anger, surprise, or even resentment towards acquaintances who firmly deny that Matthew is ‘cute’. Apprehending someone as ‘cute’ in this way may therefore not be equivalent to a mere subjective preference for this person, however intimately connected it might be in an individual’s psyche to what are merely S2-states that refer (in their mental content) to that person. There are many polysemic evaluative terms, such as ‘cute,’ which have no particular ethical significance, but whose application seems to involve an evaluative attitude with more robust interpersonal significance than brute preferences. Even when the separation seems artificial, it is important to distinguish such evaluations from connected S2 preferential desires.
We may therefore distinguish a third type of state (S3) that involves either *evaluative* attitudes or judgments of value for which reasons may be given, which anticipate agreement from others, or at least the possibility that others can understand not only that the agent has the relevant experiences, but also why the agent is motivated by these attitudes and judgments (their *illocutionary force*). Such states are sometimes described according to the medieval schema:

(S3) an appetite or desire for something (X), where (X) is apprehended in its aspect *as valuable* (or *as good*) in some respect [for reasons Y]

where the phenomenal quality of being motivated to gain or realize X may be conscious without explicit cognition of reasons Y, but such reasons are supposed to be implicit and discursively articulable on reflection (by a competent interpreter\(^{212}\)) in such a way as to draw forth agent’s agreement that this is ‘why she wanted X.’ Some S3 states may depend on more explicitly made comparative judgments (e.g. that V is *better* than W in some respect). But what is common to them all is that the evaluative attitude —the apprehension of X as good or valuable (the second half of the schema above)— usually involves the agent straightaway in a tendency to act with the intention of realizing V, unless other motivational states interrupt, override, or counteract it. In other words, the agent need not decide in some discursive thought-process to desire X because of its good aspects: rather, her desire for X flows smoothly out of her evaluation of X. In accounts that presuppose only states up to and including S3-states, decision or choice is usually construed as deliberation between X-type options that we already desire because of implicit evaluations, and it works by clarifying and comparing our evaluations so that one option in our available range emerges as better (either intrinsically, or as a means to our prior ends, or both), which makes it more ‘desirable’ as a

---

\(^{212}\) Who may be the agent herself or someone else who is expert in her evaluative tradition and its underlying theory of rationality.
result.

The distinction between these three different senses of ‘desire’ is useful in interpreting many of the issues at stake in the burgeoning literature on desire today—a literature to which I can allude here only by a few sparse examples. For instance, unsurprisingly, many anti-Humean arguments in moral psychology focus on establishing the possibility of a special kind of S3 desires that automatically follow the lead of objective value-judgments. When coupled with the prudential capacity to make reliable judgments about what contributes to flourishing, the disposition to such desires is sometimes offered as an analysis of ‘virtue.’ By contrast, defenders of Humeanism such as David Lewis characterize all desires in the S2-fashion: they are not blind urges, since they connect to beliefs about the means that will bring us what we want, but the wanting itself expresses only subjective expected value, not objective "choiceworthiness." As Lewis summarizes the Humean claim, "our actions serve our subjective expected value according to our subjective degrees of belief." On this model, "belief serves desire" only by generating an expected value for a less specific proposition A out of the values for the more specific cases that A covers or includes.

Other than these additive connections in the desire calculus, there are no necessary connections between beliefs (such as objective value-beliefs) and desires:

If there are universal correlations between certain beliefs and certain desires, that too is a contingent matter. Someone might have no desire at all for joy, knowledge, or love. Someone might believe just what you and I believe, and still have no desire at all for joy, knowledge, or love. Indeed, someone might believe just what G.E. Moore believed about the simple, non-natural properties of these things and still have no desire for them.

216Ibid, p.305.
The familiar anti-Humean who maintains the possibility of S3 desires in which the desire necessarily follows from a given evaluative belief can grant that what Lewis says is true for S2 inclinations or preferences. But in Lewis’s subsequent arguments against various anti-Humean alternatives, he seems to think that Humeanism could be false only if the role of ‘desire’ as a folk-psychological concept required that something standing in necessary connections to beliefs should count as a ‘desire.’

For example, the simplest anti-Humean theory which Lewis critiques just equates desires with certain credences or beliefs that they require—a theory that he maintains generates contradictions whenever the credences involved are less than certain. However, note that my schema for S3-type desires does not build in a necessary connection between the evaluative beliefs and the motivational state, although some conceptions of desires in the S3 sense would add this feature. For our desire D to be distinct from a preferential inclination, D only need be a state in which the agent is motivated because of her evaluation E of its objective value or choiceworthiness. This implies the counterfactual that if she lacked E, then she would not have D, but not the counterfactual that if she lacked D she would lack E. This model also leaves open the possibility that without E, she might have a quite different S2-desire (call it D2) for her object, which gives her object a high expected utility or preference-satisfaction value. When she has E, she might well have D and D2 simultaneously, and she might also have D without D2. So the presence or absence of D2 varies independently of E, and any connection they have would be contingent, since E is not what motivates a preferential inclination like D2.

\[217\] Ibid, p.306. According to Lewis, this would make the necessity a merely conceptual or definitional one, rather than "any de re necessity in nature" (p.306). Thus if you fail to care about objective value or ethical reality, the only danger is that "your inner states will fail to deserve folk-psychological names" (p.307). But that would be true only if ‘desire’ had such a definitional role in folk psychology by mere convention, rather than because our folk concepts track psychological necessities of human nature, or requirements for the flourishing of persons as such.

By contrast, evaluative desire D requires E as its counterfactually necessary condition. As noted, however, this still leaves open the possibility that we could have E without D (however ‘abnormal’ this might be): the desire’s counterfactual dependence on the evaluation does not mean that the evaluation must be a sufficient condition for the desire.

For this reason, several debates about desire today start with the understanding that something’s being desirable for an agent is distinct from his actually desiring it, and focus on various possible deliberative or communicative conditions for desirability. For example, Morton White interprets desirability as "what we ought to desire or have a moral duty to desire," and attributes to John Dewey the position that "a thing is said to be desirable just in case we know the causal antecedents and consequences of desiring it." Against this, White objects that if my desire not to smoke opium originated from considering it in a normal state of mind, and I knew the consequences of smoking it, this would hardly make the desire morally obligatory. He also holds that desirability cannot depend on our actually desiring the consequences of having the desire, and that the distinction between desire and desirability is not analogous to the distinction between apparent and real qualities. Thus a pragmatic reduction of the meaning of ‘This apple is really red’ cannot be transferred to ‘This apple is desirable.’ These are the kinds of semantic questions raised by S3-type desire, although its connected notion of objective desirability certainly need not be restricted to ‘moral desirability,’ as White seems to assume.

II.3. A Preliminary Interpretation of Desire as ‘Lack’
II.3.1. Desires vs. Evaluative Attitudes

Although much more needs to be said about the S3 category or ‘evaluative desires,’ this analysis is already enough to distinguish it from ‘pure’ evaluative attitudes. If it is possible to evaluate some object or end X as valuable or good in some respect (as healthy, beautiful, noble, etc.) without any accompanying desire for it (or aversion to it), then evaluative attitudes can be distinguished from S3-type desires or action-oriented motivations which they often accompany. Some forms of "internalism" are construed so as to deny this possibility when the S3-desire is grounded in moral considerations. But as Alfred Mele argues in a recent article, the phenomenon of "listlessness," which "consists in the total absence of motivation to engage in activities that formerly were matters of deep personal concern," is possible despite the persistence of moral beliefs or convictions. In such conditions —for example, perhaps in a serious depression— an agent may have the evaluative attitude which normally produces an S3-desire without experiencing this motivation. Aristotle also implies the possibility of such states in one of his arguments that "mind" [nous] cannot be the cause of forward "local movement:"

Further, neither can the calculative faculty or what is called ‘mind’ [nous] be the cause of such movement; for mind as speculative never thinks what is practicable, it never says anything about an object to be avoided or pursued, while this movement is always in something which is avoiding or pursuing an object. No, not even when it is aware of such an object does it at once enjoin pursuit or avoidance of it; e.g. the mind often thinks of something terrifying or pleasant without enjoining the emotion of fear. It is the heart that is moved (or in the case of a pleasant object some other part). Instead, Aristotle says, "mind practical" is "capable of originating local movement," but only by calculating means to ends given by appetite (in the general sense described above), "for

---

223 Aristotle, De Anima, III.9, p.597, 432b 27-34.
that which is the object of appetite is the stimulant of mind practical." 224

On this model, however, we can have evaluative attitudes without accompanying desires, emotions or "passions," but only if they are purely speculative judgments. But if "evaluative attitudes" are to constitute a separate category in their own right, they must somehow be distinct from motivationally neutral, pure cognitions made from a third-personal perspective as well. At minimum, to constitute an "attitude" rather than a particular judgment, the evaluative attitude has to consist of a sustained disposition to make relevant (objective and communicable) evaluations of the right kinds of objects over time. For example, even if it has additional emotional elements, person A’s trust in person B must involve a disposition to evaluate the motives of B’s actions as loyal to A, or at least as taking A’s interests into account, even when the ‘preponderance of evidence’ as assessed neutrally by a third party would suggest otherwise. 225 This does not require naive blindness over many years, nor refusal to recognize evidence beyond any reasonable doubt that one has been betrayed, but it does involve an attitude that, to stick with my juridical metaphor here, always decidedly slants the ‘burden of proof’ in favor of B. Similarly, to take a more contentious example, enlightened patriotism (as opposed to undiscriminating enthusiasm for one’s nation and its symbols) may consist in a kind of public analog to trust, namely an evaluative attitude which gives one’s nation and its government some benefit of the doubt, because it has earned respect over many generations and because its institutions require public loyalty in order to function.

We have not considered the relation between evaluative attitudes and emotions, but it is worth asking whether the former could be explained as patterns of evaluative judgment that


225In this regard, see C. S. Lewis’s discussion of faith in "On Obstinacy in Belief," Screwtape Proposes a Toast and other Pieces (Collins/Fount Paperback, 1977): 59-74.
are both distinct from desires they foster and intelligible only from the perspective of personal involvement in relationships, practices, and institutions—i.e. those social contexts in which the values they are about acquire something analogous to the objective status of facts. On this hypothesis, evaluative attitudes will be "backward-looking" in Elizabeth Anderson’s sense: the subject with an attitude will be disposed to make her evaluations on the basis of whether it "bears an appropriate relation of narrative unity to prior actions" and evaluations.\textsuperscript{226} This hypothesis would help explain why evaluative attitudes, unlike the neutral judgments of speculative intellect, are closely tied to what makes sense from the perspective of the subject’s involvement with others in narrower and wider interpersonal contexts. This suggests how evaluative attitudes can retain their objective intentionality while remaining distinct in structure from the purely cognitive judgements which are the ideal of science. As Jorge Arregui writes in an interesting paper on emotions—notably, after citing C.S. Lewis—"evaluation depends necessarily on the subject, but what is evaluated is the object, and therefore the evaluative property is attributed to the object and not the subject."\textsuperscript{227} Evaluation is perhaps a distinct mode of intentionality which picks out axial aspects of the object made salient by social relations, practices, and meaning-contexts, which are different than the aspects thought in neutral speculative cognition.

\textbf{II.3.2. Plato’s Analysis of Desire}

In this analysis, I will not focus on emotions (S5) and evaluative attitudes (S6), however, since their clarification depends on first developing a better understanding of ‘desire’ by

\textsuperscript{226}See Elizabeth Anderson, "Reasons, Attitudes, and Values: Replies to Sturgeon and Piper," \textit{Ethics}, Vol. 106, No.3 (April, 1996): 538-554. This essay was part of a symposium on Anderson’s book, \textit{Value in Ethics and Economics} (Harvard University Press, 1993). Anderson’s concern in this passage is with the rationality of action, but I have applied her idea here to evaluative attitudes.

reflecting on the three kinds that were provisionally distinguished above: impulses (S1), preferential inclinations (S2), and evaluative desires (S3). What justifies calling them all states of ‘desire,’ despite their very significant differences in cognitive structure, is a general feature of the _motivation_ the agent experiences in any state of these three types, which we might call its _teleological orientation_: the agent is ‘moved’ towards the end of satisfying the desire. This is true even in the case of brute impulses, although as moved by them we lack any concrete conception of their end, which remains generic.\(^{228}\) In contemporary literature, this teleological feature is so familiar that it is assumed without question to be the form of motivation _as such_, and so it hardly ever becomes a theme for reflection. But in ancient moral psychology, Plato first attended to this feature in two well-known discussions of desires as experiences of ‘lack’ that anticipate objects which would make up for what is lacking in the subject who has the desire. The first occurs near the beginning of Socrates’s argument for a tripartite division of the soul in the _Republic_

> What then, said I, of thirst and hunger and the appetites generally, and again, consenting and willing...? Will you not say, for example, that the soul of one who desires either strives for that which he desires or draws towards its embrace what it wishes to accrue to it, or again, insofar as it wills that anything be presented to it, nods assent to itself thereon as if someone put the question, striving towards its attainment?\(^{229}\)

Socrates classifies ‘will’ (wish and choice) with desire here, but for the moment I am

\(^{228}\)It is wrong to think that an end must be explicitly conceived or even conceivable to count as the _telos_ of some motivational state. For in Thomistic accounts of final causation, happiness or eudaimonia has a sort of generality _similar_ to that of S1 states in my account. Although the will’s true telos is the divine goodness which alone can fulfill our desire for happiness, it exercises its final causation even on the agent who takes lesser or apparent goods as his explicit last end: even evil ultimately derives all its motivation from the primal urge towards a general good whose attraction it misinterprets in the false vessel of a lesser worldly image of the good. Thus the last end itself is not clearly apprehended, despite its motivational pull.

\(^{229}\)Plato, _Republic_ IV, 437b-c, p.679. The Rouse translation reads: "...thirst and hunger and desires in general, and again to wish and to be willing...the soul of the desirer always wants that which he desires, or is attracted to that which he wishes to have; or again, inasmuch as it wants something to be provided for him, it nods ‘yes’ to itself as is someone had asked the question, reaching forward to the production of the thing. Will you not agree?" See _Great Dialogues of Plato_, tr. W.H.D. Rouse (Mentor/Penguin, 1984), p.236.
concerned only with the basic feature of motivation he detects: namely, the intentional structure of desire. Each desire has an intentional correlate that is essential to its identity: "Each desire in itself is of that thing only of which it is its nature to be."\textsuperscript{230} From this it follows that "mere thirst" has "mere drink" as its object, whereas only a qualified version of this basic desire (such as ‘hot thirst’) desires ‘cold drink’ as its object.\textsuperscript{231} Although this distinction between pure and qualified desires and their correlates has seemed contentious to some commentators concerned about its role in the argument for the division of reason and desire, it is viable if understood in terms of my earlier distinction between S1 and S2 states. A raw urge has an intentional object, but it is generic and without any cognitive specification, whereas preferences are qualified versions of such urges, directed onto conceptualized objects. Socrates also insists that a desire for "good drink" rather than for drink as such is a qualified desire, distinct from the impulse.\textsuperscript{232} Since good refers to an objective evaluation, this would be an S3-desire for some liquid because it was judged to be good in some cognitive sense, i.e. a judgment that makes a claim on others. By contrast, "thirst that is just thirst is neither of much nor little nor good nor bad, nor in a word of any kind..."\textsuperscript{233} Qualifications of the desire-object by kind-concepts take us into states beyond the S1 level, such as preferential inclinations or evaluative desires.

Of course, in the \textit{Meno}, Plato famously had Socrates argue that people can only desire what appears ‘good’ to them (in the sense that it brings advantage or contributes to their happiness), which seems to imply that some rational evaluation, whether right or faulty, is

\textsuperscript{230}Republic IV, 4373, p.679.
\textsuperscript{231}Republic, IV, 437d-e, p.679.
\textsuperscript{232}Republic, IV, 438a-b, p.679-80.
\textsuperscript{233}Republic, IV, 439a, p.681, my italics.
implicit in every motivational state.234 Relative to this simplistic analysis, the moral psychology of the Republic seems to be a clear advance, since it distinguishes rational evaluation from appetite, and thus allows for motivational states which intend their objects without any judgment that they are ‘good’ in any objective sense. Thus as Gary Watson argues in a famous paper on freedom, there is a real "distinction between wanting and valuing," which is central to Plato’s ethics.235 While "On Hume’s account, Reason is not a source of motivation," for Plato,

...the rational part of the soul is not some kind of inference mechanism. It is itself a source of motivation. In general form, the desires of Reason are desires for ‘the Good.’236

Rational desires, in other words, have the structure of S3-motivation defined above. On Watson’s reading, for Plato there is a necessary connection between the evaluation and the motivation: "to think a thing good is at the same time to desire it (or its promotion),” and thus the values it confers provide "reasons for action." By contrast, lower desires have objects that "may not be thought good."237 Although this contrast tends to suggest that there is no state of between raw sensitive impulses (S1) and boulesis or evaluative desire (S3), the advance on the Meno account is still clear.

In another way, however, the Republic analysis preserves a central part of the intuition at the work in the Meno account, namely the idea that desires of any sort —pure or qualified— are ‘attractions’ towards something which, insofar as it ‘attracts’ functions as a ‘good.’ In a nonjudgmental sense, then, mere food is the ‘good’ for which mere hunger is the desire. The

---

234Plato, Meno, 77c-78b.
236Ibid, p.98.
Republic recasts the Meno principle that ‘every desire is for what is evaluated or represented as good’ into the more basic principle that every desire is for its ‘good’ or moved towards the end that will satisfy it. This principle is more inclusive in application, since it covers S1 and S2 states as well as S3 evaluative desires. Thus although Plato emphasizes the contrast which Watson highlights between ‘valuing’ and ‘desiring’ —or more accurately between rational appetite and lower desires— Plato also recognizes that they have something in common: every desire has the motivational structure of a yearning for satisfaction. Thus “The soul of the thirsty, in so far as it thirst, wishes nothing else than to drink, and yearns for this and its impulse is towards this.”238 In general, "affections," like "diseases," produce impulses that "draw and drag" the agent towards their end, just as do positive valuations by the rational part of the soul.

One might be tempted to object that Plato meant this analysis to apply only to lower, instinctive appetites such as hunger and thirst, since he immediately goes on in Republic IV to analyze "high spirit" and reason in different terms, and S3-desires at least require some involvement of practical cognition. But as we see in the Symposium, in Socrates’s and Diotima’s speeches, the ‘lack-structure’ of desire is explicitly reformulated in order to extend it to motivation by the good, i.e. or to what I have called S3-type states involving objective evaluation of the conditions of our well-being. Socrates begins with the point that love is intentional, since "it is the nature of Love to be the love of somebody..."239 Moreover, its attitude towards its object is one of longing:

Then isn’t it probable, said Socrates, or rather isn’t it certain that everything longs for what it lacks, and that nothing longs for what it doesn’t lack?

This implies, contra Agathon’s speech, that love lacks the divine beauty and happiness for

238 Republic, IV, 439a-b, p.681.
239 Plato, Symposium, 199d, p.551.
which it longs. This rule seems to ignore that we sometimes continue to long for good things such as health and riches when we already have them. But against these potential counterexamples, Socrates notes that in such cases, "what you want is to go on having them, for at the moment you’ve got them whether you want them or not.” And since the contingency of fortune and mortal fate deprives us of the certainty of enjoying such things permanently, "desiring to secure something to oneself forever may be described as loving something which is not yet to hand.”240 We may add here a further point here to supplement Plato’s analysis. If someone objects that we must have a desire to possess good things in the present moment in order to explain our pleasure or happiness in having them right now, the rule that we can only desire what we lack can be made counterfactual: to desire X entails that if one lacked X, one would long for it and be in that respect dissatisfied without it. Thus our satisfaction with X in the present moment of possessing it can be explained by our sense that we have avoided the desire we would experience in the nearby possible worlds in which we lack X. Our happiness is explained not by our desiring X simultaneously with having X, but rather by the (temporary) termination of the desire by the presence of X in our possession. The upshot of this analysis is Plato’s most general formula for desire:

And therefore, whoever feels a want is wanting something which is not yet to hand, and the object of his love and of his desire is whatever he isn’t, or whatever he hasn’t got—that is to say, whatever he is lacking in.241

And from this it easily follows that love itself cannot be the beauty which it desires. This general formula needs refinement, however, because as it stands it would imply that we desire anything we happen to lack, but only some of these ‘absences’ are salient, because only some of them become intentional objects. Diotima’s speech accomplishes this

240Ibid, 200d-e, p.552.
241Ibid, 200e, p.553.
refinement.

Diotima begins by fending off an objection to the conclusion of the argument against Agathon: namely that it would entail that Love is "bad and ugly." Instead, she says that the ‘lack’ implied by desire does not entail the opposite quality, but is an intermediate state: to desire beauty is to be ‘on the way’ to it, rather than simply to lack it, like someone who is ugly. Similarly, she portrays love as "halfway between mortal and immortal," a "powerful spirit" that acts as medium or hermeneut between impoverished human beings and the gods who enjoy perfect beauty and goodness. In her invented fable, Love is the son of Need and Resource, begotten on the same day as Aphrodite’s birth. The implication is that Love has enough resources at least to conceive what it lacks and thus to start on the way towards it or to pursue it—just as in the theory of ‘recollection’ or anamnesis, some trace of the knowledge lacked is requisite in order to make its absence salient and make learning possible. Thus Diotima adds that just as the gods do not seek for wisdom and truth, since they have it,

Nor, for that matter, do the [totally] ignorant seek the truth or crave to be made wise. And indeed, what makes their case so hopeless is that, having neither beauty, nor goodness, nor intelligence, they are satisfied with what they are, and do not long for the virtues they have never missed.

The ‘lack’ which is essential to desire therefore cannot be a complete lack, or one that seems irrelevant to us, i.e. an ‘absence’ to which we are indifferent. Rather, it is one that is felt as a void precisely because we have some partial sense of the end, some anticipatory trace or hint of its importance for well-being, which allows it to become our intentional object, or to

---

242Ibid, 202b-d, pp. 554-5.
244And Diotima’s analogy here nicely squares with the image of Love as a trickster-figure: "he brings his father’s resourcefulness to his designs upon the beautiful and the good, for he is gallant, impetuous, and energetic, a mighty hunter, master of device and artifice—at once desirous and full of wisdom..." (203d).
245Ibid, 204a, p.556, my italics.
‘hook onto’ us, drawing us into the intermediate state of ‘becoming’ or moving towards it. Being able to desire higher goods is thus itself an advantage over creatures that cannot even experience their lack in the requisite way.

Diotima then argues that objects gain their relevance to us —and thus make their absence salient and motivate ‘desire’— by their apparent goodness. This argument commences after the dialectical turning point when Socrates asks what value Love can have for us if it is not itself the good or beautiful. Diotima responds indirectly, beginning with the suggestion that desire for the beautiful is really only one part or aspect of Love: "For ‘Love, that renowned and all-beguiling power,’ includes every kind of longing for happiness and for the good." In fact, she insists that love properly includes only desire for what is evaluated as beneficial:

I know it has been suggested, she continued, that lovers are people who are looking for their other halves. But as I see it, Socrates, Love never longs for either the half or the whole of anything except the good. For men will even have their hands and feet cut off if they are once convinced that those members are bad for them.

The desire to lose a limb evaluated as fatally diseased is an S3-desire, just like the desire not to drink bad water. In the Republic, of course, the thirsty man’s aversion to drink he evaluates as poisoned was attributed to reason, and ‘desire’ as a technical term for the lowest part of the soul was restricted to S1 urges (and possibly their S2 applications). But we see in the Symposium that S3 motivational states share the same general lack-structure —indeed, they display it more clearly and definitely. Love, in the most general sense of a movement drawn by the absence of something ‘good’ as intentional object, is thus found at all three ‘levels’ of the tripartite soul, from the instinctual desire of beautiful bodies, to the ‘spirited’

---

246Ibid, 204c, p.556.
247Ibid, 205d, p.557.
248Ibid, 205e, p.558.
desire for honor and for noble "laws and institutions," to the intellectual loves of scientific knowledge, philosophical wisdom, and finally the Form of Beauty itself. On this analysis, as Diotima says, "Love is not exactly a longing for the beautiful," as Socrates has said, because it longs instead for "the conception and generation that the beautiful effects," since it is these which are qualified as good. Desire is directly for the apprehended good (the product), and thus only indirectly for the beauty which facilitates the product. Even sexual desire is at bottom an existential desire for the good of immortality, which we can achieve only deficiently through procreation. The same end would seem to be the real ‘good’ for the sake of which the Form of Beauty itself is desired, since genuine immortality of the soul is said to be the product of this Form’s revelation: "if ever it is given to mortal man to put on immortality, it shall be given to him."

—Gary Watson’s Analysis

One might object to this analysis that it implies that all desires give the agent at bottom the same sort of reason to act, namely to satisfy the desire. As Gary Watson says any desire may provide the basis for a reason in so far as the non-satisfaction of the desire causes suffering and hinders the pursuit of ends of the agent. But it is important to notice that the reason generated in this way is a reason for getting rid of the desire, and one may get rid of a desire either by satisfying it or by eliminating it in some other manner (by tranquilizers, or cold showers). Hence this kind of reason differs importantly from the reasons based upon the evaluation of the activities or states of affairs in question. For in the former case, attaining the object of desire is simply a means of eliminating discomfort or agitation, whereas in the latter, that attainment is the end itself.

249 Ibid, 210c, p.562.
251 Ibid, 206e-207d, p.559.
252 Ibid, 212a, p.563.
Our prior distinction between S1-S2 states and S3 valuations is sufficient to accommodate Watson’s point: in S3-type desires, it is the reason for valuing an activity or state of affairs that motivates desire for it, whereas in S1-S2 desires, it is the subjective attractiveness of the object that motivates. Thus Watson is wrong to say that the difference between lower desires and valuing is all in their source rather than their content.\footnote{Ibid, p.102.} as we have seen, there is a real difference in the aspects picked out in their intentional contents, because S3 desires have objective evaluative content (the object-as-good-for-so and so-reasons) and S1 and S2 desires do not. He is right that lower desires also generate a kind of minimal, disjunctive ‘reason’ either to satisfy them or get rid of them, but this reason to act on them is not what actually motivates the agent’s desire for the object in real S1 or S2 states: rather, it is the object’s subjective attractiveness which does that work. In S3 states, a different, non-disjunctive reason to attain the end is what motives the desire for the object, and when we are motivated in this way, as Watson insightfully points out, "We aim to satisfy, not just eliminate, [rational] desire."\footnote{Ibid, p.101.} But similarly, in S1-S2 states, the end’s attractiveness generates is a motivation to get satisfaction by possessing the object, not just by eliminating the desire. Watson’s argument thus poses no obstacle to my comprehensive lack-analysis of desire, for this analysis recognizes the significant differences in motivation involved in S3 as opposed to S1 and S2 desires, while still identifying the same basic teleological structure, direction, or formal dynamic of motivation which is operative in all types of desire.

On my interpretation, Plato recognizes something like this: for him, achieving the goal of any desire means restoring a kind of motivational equilibrium or harmony which the salience of the lacking object has disturbed, in S3 states just as in lower appetitive S1-S2 states. This
is not to say that this ‘equilibrium’ or ‘requiting the felt incompleteness’ itself is the end to which the object of desire is a mere means, but rather that the character of the motivation is always such that its intentional correlate appears under the aspect of an ‘object that fulfills a lack,’ however else this content and the kind of ‘lack’ it involves may differ. For example, I have an S2 inclination for ‘a beer-as-quenching-thirst’ or I have an S3 desire for ‘education-as-good-because-it-imparts-knowledge and... which-I-lack.’ All desires are towards objects intended as filling salient gaps in some aspect of the agent’s overall happiness or equilibrium: that the object is anticipated to fulfill a lack always modifies the intentional content of the object of any desire, whether S1-S3. This is Plato’s insight, though he took it to be a fact about motivation in general, rather than simply about desiderative motivation.

II.3.3. Some Further Points From Aristotle

Against the background of this analysis of desire on the teleological model of eros, it is easier to understand why Aristotle insists on the unity of "appetite" (orexis) as one general power of the soul with diverse manifestations. As we have already seen, Aristotle repeatedly links lower "desires" (epithumia) to sensation and immediate physical gratification: if an animal has "sensation, then necessarily also imagination and appetition, for where there is sensation, there is also pleasure and pain, and where these, necessarily also desire." Hunger and thirst are typical instances of desire in this sense, and since desires arise in response to "feelings of pleasure and pain," they are not influenced by long-run considerations, but only by "what is just at hand: a pleasant object which is just at hand

---

256 Aristotle, De Anima, II.2, p.558, 413b 21-23.
257 Ibid, II.3, p.559, 414b 12.
258 Ibid, III.11, p.599, 434a 1-3: "Clearly they have feelings of pleasure and pain, and if they have these, then they must have desire." So these feelings are sufficient conditions for pain.
presents itself as pleasant and good, without condition in either case, because of want of foresight... Desire in this sense seems to be ambiguous between what we have described as raw biological urges without representation (S1) and inclinations that are directed to preferred objects (S2).

To link Aristotle’s conception of desire to our previous analysis, it will help to further subdivide the S2 category into those desires that arise directly from the specification of instinctive impulses by intending particular objects represented as satisfying these impulses, and other "preferences" which similarly have intentional content without objective evaluation, but are not simple specifications of generalized urges. In both cases, instinct alone is insufficient to explain the S2-desire, since there are other objects that might satisfy our instinctive urges besides the objects we actually desires. But an S2 desire for chocolate ice cream represents something as pleasant in a way more closely linked to satisfaction of a biological urge than, say, a desire to read biographies (as opposed to romance novels). Yet the latter stands on the S2 level if it involves no evaluation that makes a claim on others, or invokes no principle with interpersonal significance. Let us call the former ‘biologically grounded inclinations,’ and the latter ‘culturally grounded inclinations.’

In his attempt to distinguish desires from ‘valuations,’ Gary Watson makes a point relevant to this subdivision of the S2-category. He notes that desires other than "appetitive or passionate desires" (e.g. those for a particular kind of food or sexual partner) may exhibit "independence of evaluation," meaning that we can experience them ‘against our will’ or

---


260This is not to deny that ‘cultural’ factors in the most general sense will usually play a role in how exactly an instinctive urge is specified, or what objects we fix on, since S2-inclinations are open to a large amount of adaptation and social molding, as the influence of television advertising shows (even the desire for chocolate ice cream may be influenced by seeing others we admire preferring it to vanilla). Moreover, I do not deny that there will be borderline cases between the clear domains of these two subdivisions.

‘unfreely.’ For example,

One may be disinclined to move away from one’s family, the thought of doing so being accompanied with compunction; yet this disinclination may rest solely upon acculturation rather than upon a current judgment of what one is to do....Or, taking another example, one may have been habituated to think that divorce is to be avoided in all cases, even though one sees no justification for maintaining one’s marriage. In both of these cases, the attitude has its basis solely in acculturation and exists independently of the agent’s judgment.262

These aversions as Watson construes them are culturally-grounded inclinations, rather than S3 desires; even though they seem "more akin to evaluation than to appetite," since they may be expressed in evaluative language,263 they are only habitual preferences lacking the strong evaluation involved in S3-desires. There are presumably many similar preferences and inclinations which have cognitive content without objective evaluative significance, and yet remain only distantly related to innate biological instincts and the impulses they generate.

On this analysis, Aristotle’s sense of "desire" as epitumia lines up with biologically grounded inclinations, while his notion of "passions" seems closer to culturally grounded preferences and the emotions that go with them. His conception of desire emphasizes what such S2-inclinations have in common with raw S1 states: namely, an intimate connections to bodily needs, biological instincts, and physiological alterations produced by physical stimuli.

Though S2 inclinations on my analysis need not be motivated ultimately by the pursuit or physical pleasure or avoidance of physical pain, in one subcategory of them the connection will be very clear. Desire in this sense, however, is only one of three species of "appetite" for Aristotle:

If any order of living things has the sensory, it must also have the appetitive [orektikon]; for appetite is the genus of which desire [epithumia], passion [...] and, wish [boulesis] are the species; now all animals have one sense at least, viz. touch, and whatever has a sense has a capacity for pleasure and pain, and therefore has

263Ibid.
pleasant and painful objects present to it, and wherever these are present, there is
desire, for desire is just appetition of what is pleasant.\textsuperscript{264}

We may think of "appetite" as Aristotle’s general term for any state of \textit{action-oriented}
motivation, which is always aimed at some object as its end. Notably, the three kinds of
motivation he distinguishes align with the parts of Plato’s tripartite soul. In his critique of
thinkers who divide the soul only into the "calculative, the passionate, and the desiderative"
(or desiring), Aristotle takes it as a reductio of their position that they cannot treat appetite in
general as one of the basic faculties:

It is absurd to break up the last-mentioned faculty, as these thinkers do, for wish is
to be found in the calculative part and desire and passion in the irrational, and if the
soul is tripartite appetite will be found in all three parts.\textsuperscript{265}

The implication is that "wish" is motivation produced by the rational part of the soul, while
"passion" is the motivation produced by the spirited part, and desire the motivation produced
by the lowest part of the tripartite soul. Aristotle holds that it is more relevant to treat
appetite as a unified faculty (the \textit{orektikon}) because otherwise it would cut across the
tripartite soul, leaving each of Plato’s three parts with both motivational \textit{and} non-
motivational aspects, which seems peculiar. On the tripartite account, mind would then
engage in speculative thinking, but also in wishing, and the lowest part would engage in
sensation, but also in desiring. To Aristotle, it makes more sense to treat mind, sensation,
and appetite in general as the basic types of soul, thus allowing \textit{all} forms of motivation to be
rooted in a single faculty:

That which moves therefore is a single faculty and the faculty of appetite...As it is,
mind is never found producing movement without appetite (for wish is a form of
appetite; and when movement is produced according to calculation it is also
according to wish), but appetite can originate movement contrary to calculation, for
desire [\textit{epithumia}] is a form of appetite. Now mind is always right, but appetite and

\textsuperscript{264} Aristotle, \textit{De Anima}, II.3, p.559, 414b 1-5.

\textsuperscript{265} Ibid, III.9, p.596, 432b 5-8.
imagination may be either right or wrong. That is why, though in any case it is the
object of appetite which originates movement, this object may be either the real or
the apparent good.\textsuperscript{266}

The tripartist could respond that it is notable that appetite divides into \textit{three} forms
corresponding to his three parts. Commenting on this passage, Lear says,

...we find wishes in the part of the soul which reasons about how to act, and we find
desires in the ‘irrational’ part of the soul: for example, the basic appetites for food
and sex....It seems that either Aristotle must give up the idea that the soul has parts,
or he must find a way of conceiving the source of movement to be a single part of
the soul. He chooses the latter option. There appear, he says, to be two sources of
movement, practical mind and appetite. Practical mind differs from theoretical
mind in that it is concerned with how a desire can be satisfied....Aristotle located
both practical mind and appetite within a single faculty of the soul responsible for
movement: the desiring part [\textit{to orektikon}].\textsuperscript{267}

Thus Lear, like most Aristotle scholars, portrays Aristotle as holding a fundamentally
bipartite model of the human soul, in this form:

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
\textbf{The Appetitive soul [orektikon]} & \textbf{The rational soul} \\
\hline
\textit{desires} & \textit{wishes} & \textit{Deliberate choice} \\
\textit{passions} & \textit{boulesis} & \textit{Speculative mind} \\
\textit{(epithumia)} & \textit{(prohairesis)} & \textit{(nous)} \\
\hline
\end{tabular}
\end{center}

But I am not sure that the text just cited quite supports this construal. It does suggest that
wish can be for an object that appears to be but is not actually good, implying that wish is
not strictly speaking part of reason or \textit{nous}. But this may not mean that ‘practical mind’ is
simply part of appetite, but rather that wish is the aspect of appetite that connects mind to
motivation, which \textit{combine} to make ‘practical mind.’ Thus in his \textit{Nicomachean Ethics},
Aristotle says that the "irrational element" of the soul includes both nutritive and appetitive

\textsuperscript{266}Ibid, III, 10, p.598, 433a 22-27.

\textsuperscript{267}Lear, \textit{Aristotle}, p.143. Note that he means by "desiring" what our translation means by "appetitive" (and
in note 113 Lear critiques the "appetite" translation as misleading).
divisions, and the latter (or at least part of it) "shares in a rational principle"\textsuperscript{268} since in continent, temperate, and brave men it obeys or "is in some sense persuaded by reason."\textsuperscript{269} If this is right, practical mind or ‘intellectual appetite’ (as it is often called) does seem to form a kind of ‘middle soul’ between the strictly rational and strictly irrational in Aristotle’s conception, although it may be a compound middle soul, unlike Plato’s thumos.

There are contested issues here for Aristotle scholarship, but what is interesting for my purposes in Aristotle’s approach is his attempt to find one universal structure for all motivation on the model initially furnished in Plato’s \textit{Meno}: namely, the teleological schema in which movement or action is the result of the agent’s being drawn by an object apprehended as ‘good’ in some way —real or apparent. His understanding of "wish" (\textit{boulesis}) or the motivation produced in "practical mind" is obviously similar to the schema for desire in the S3-sense distinguished above.\textsuperscript{270} But "desires," or biological inclinations in the form of S1 or S2 states, are supposed to have a similar structure: the object appears as pleasant first, and the agent is drawn towards it like a beacon second (in logical, though not necessarily temporal order). Thus while "wishes" have real evaluative contents (as in my S3 schema), and "desires" are mere S2 states with representational content but not evaluative interpersonal significance, Aristotle says they are linked by the fact that in both cases, the object of appetite starts a movement and as a result of that thought gives rise to movement, the object of appetite being to it a source of stimulation.\textsuperscript{271}

Aristotle’s insight is that what both S2 and S3 states have in common is their appetitive


\textsuperscript{269}Ibid, 1102b.26-31, p.1742.

\textsuperscript{270}And as Lear explains, in the \textit{Nicomachean Ethics} "deliberated choice" (\textit{prohairesis}) or decision follows "only after a process of deliberation. Aristotle’s theory of deliberation (\textit{bouleusis}) is a theory of the transmission of desire. The agent begins with a desire or wish (\textit{boulēsis}) for an object. The object of wish appears to be good to the agent. But the appearance helps to constitute the wish itself...an agent’s awareness that he wishes for a certain end is itself a manifestation of that wish" (Lear, \textit{Aristotle}, p.143).

\textsuperscript{271}Ibid, III.10, p.598, 433a 18-20, my emphasis.
form, or in Platonic terms, their basic ‘erotic’ structure of feeling a lack or incompleteness in the presence of an object that seems to promise completion, or satisfaction of the lack. As Lear says (using the term "desire" widely for orexis or ‘appetite’ in general),

> There must be something which moves animals to move, and this motive force is desire. Desire and animal movement have a similar structure: desire is desire for an object which the animal is lacking, an animal movement is directed toward the object of desire...Humans distinguish themselves from other animals by their ability to think and by the fact that in addition to appetites [epithumia or ‘desires’ in our translation] they have more sophisticated desires—for example, the desire to understand. Human action cannot be understood merely as an attempt to satisfy basic appetites.\(^{272}\)

This lack-structure shared by all the different species of appetite is the reason why, as Lear emphasizes, awareness of an appetite is an essential part of the motivational state itself.\(^{273}\)

Thus, like Plato’s, Aristotle’s texts support the idea that although they differ in the status of their cognitive content, preferential inclinations and evaluative desires share two important features: (a) the attraction of the object as represented or intended is primary, and the subjective feeling of the agent is secondary or a response to the object, and (b) the object is inclined towards or desired because realizing it not only contributes to well-being, but does so in the particular sense of fulfilling a want, or tending to restore the agent to a kind of equilibrium of which the desire itself is a disturbance (or ‘affection,’ in the old sense).\(^{274}\)

These two features make up what I call the fundamental structure of ‘desire-as-lack.’

On this basis, we can recognize that a ‘desire’ for X in any mode involves a sense on the part of its agent, however vague and perhaps unarticulated, that accomplishing or gaining X will requite an incompleteness or restore a lost balance. Although the type of ‘satisfaction’

\(^{272}\)Lear, Aristotle, p.142, my italics for emphasis.

\(^{273}\)Ibid, p.149.

\(^{274}\)Some readers may find my attribution of this view to Aristotle as unjustified, because in the Nicomachean Ethics and other ethical writings, he develops a notion of motivation for the sake of virtue alone that is ostensibly connected with the agent’s highest good but is not as egoistic as my formulation of S3-desires implies. This topic is taken up subsequently here in §II.5.3.
this involves differs quite substantially between S1-S3 (ranging all the way from brute urge-gratification, through satisfaction of a subjective preference, to happiness in having gained something that is good or valuable for life in an objective sense), in all these cases the agent’s original ‘upsurge’ towards the goal, or first-personal motivation to act, is a response to this sense of ‘lack;’ it is an anticipation of filling in an absence. S1-S3 states of motivation share the structure of a ‘pull’ on the agent; this pull grasps the agent, who is initially passive, and his or her motivation initially arises only in response to this pull. In dynamic terms, ‘end-pull’ rather than ‘agent-push’ is primary in desires of all types. In S3 states, of course, the desire is a response to rational grounds for wanting its object, or to apprehensions of objects and ends as valuable for life in different ways for various (implicit) reasons, but it is still a response to these perceptions and judgments —as we see paradigmatically in St. Thomas’s theory of the dependence of appetite on evaluative intellect. For example, in discussing God and the Holy Spirit, Aquinas argues that every "intellectual nature" must have a will, because understanding must incline it to its proper end (in a fashion analogous to natural or nonvolitional appetites):

...every inclination of the will arises from this: by an intelligible form a thing is apprehended as suitable or affective. To be affected toward something—so far as it is of this kind—is to love that thing. Therefore, every inclination of the will and even of the sensible appetite has its origin from love. For from the fact that we love something, we desire that thing if it be absent...275

This is a clear statement of the idea that all motivation (including intellectual motivation, or volition) is desiderative in the general sense of Platonic eros or teleological impetus.

II.4. Nine Types of Second-Order Desire

275 Thomas Aquinas, Summa Contra Gentiles, Book IV, ch. 19.3, p.117. I am indebted to Dan Moloney for this reference.
This overview of different appetitive states of motivation provides a basis for considering another important question that Plato, Aristotle, and their successors did not address: in what senses, if any, might it make sense to say that there can be second-order desires, or desires to have certain first-order desires (i.e. desires whose ends are not other motivational states). This question will have important ramifications both for our analysis of the will and volitional identification in Chapters IV and V. The foregoing distinction between impulses (S1), inclinations and preferences (S2), and evaluative desires (S3) suggests nine logically conceivable combinations: S1-S1, S1-S2, S1-S3; S2-S1, S2-S2, S2-S3; S3-S1, S3-S2, and S3-S3. I will briefly explore each of these conceivable permutations in turn.

(1) We can immediately dismiss the idea of an S1-impulse for another impulse, since S1-states in general derive from instinct, physiological need, or deformations of these. But not even in any known pathology is there such a thing as a conscious urge (without particular cognitive content) to experience some other conscious urge. We cannot even imagine what it would be like, for example, to experience a craving to feel thirsty, since none of the urges or impulses with which we are familiar could be ones it would make sense to have an urge or drive to experience. I might wonder or be curious what it is like for a bird to experience a drive to fly south, or the salmon to return upriver, but I can hardly feel an urge or impulse to experience their general type of drive-experience. Similarly, a person who has been deprived of some kinds of feeling that are natural to us (say, sight or touch), will miss these experiences and perhaps even feel a longing to fill in this gap in their experience. But this is still not a bare urge to experience an impulsive state of motivation. We cannot really imagine a craving for a craving.

(2) But if an S1-impulse for an S1-impulse is impossible, then so is an S1-impulse for and S2-inclination or preference. It makes no sense to have an impulse to acquire
preferences or inclinations of some general kind, since these by definition include cognitive content, whereas the impulse does not. I cannot have a non-cognitive desire whose telos or end is a motivational state involving cognition and intentionality. For the impulse to aim at any general kind of such motivational states, it would have to have cognitive content itself. I cannot experience thirst to desire a beer, nor can I crave to desire chocolate ice cream—such combinations make no sense. I can remember with fondness my childhood desire to build with legos, and perhaps wish to live it again, but I cannot have a non-cognitive impulse or drive towards such a specific preference.

(3) For the same reasons, there is even less real possibility of S1-impulses for S3-evaluative desires. It is conceivable that one could experience an impulse to think, to use one’s evaluative capacities, to develop the ability to shape one’s motivations by critical reflection, because we often experience urges to use or test out our natural powers. But this does not amount to an impulse for any particular S3-desire, or any general category or them. The appeal we might see in conforming one’s desires to one’s evaluative judgments is not a basic urge to do this, however it is described. So in sum, it does not seem that S1-states could ever be second-order, or have any other type of first-order desire as their end.

(4) Prospects are better for S2 inclinations and preferences. Yet it is difficult to imagine an S2-inclination to experience, for no objectively articulable reason, an S1-impulse. Harry Frankfurt discusses the case of a doctor who is trying to help narcotics addicts, and believes he might better be able to help them if he could experience "what it is like for them to desire the drug to which they are addicted." Since the desire for a particular drug, say the desire to take cocaine, would be an S2 inclination, we need to modify the example, and imagine that what the doctor wants to experience is the general urge for a high that addicts longing

for several different types of drug experience. Frankfurt argues that since his doctor only wants to experience the addict’s desire, without wanting it to be effective in moving him to act on it and to take the drug, he "now wants only to *want* to take it, and not to *take* it," and may even arrange to make it impossible to satisfy the urge once it is induced.  

Such a case makes sense because the agent’s sense of lack or anticipation of rectifying a felt incompleteness, which is essential to any desire, is *not transitive*: to desire to want a drug is to feel less well off *without necessarily* experiencing the incompleteness that an addict experiences. The doctor feels worse off because he cannot empathize with his patients as well as he could if he could temporarily undergo their condition, but he does not miss the pleasure of the cocaine rush. So motivation via his second-order sense of ‘lack’ does not entail that he presently feels incomplete or any less satisfied without getting the high itself —the high for which he *would* feel a need if he experienced the addict’s S1-impulse (in our modified case). In general, a genuinely second-order desire D2 (of any kind) for another desire D1 (of any kind) requires that having D2 not constitute having D1 —otherwise D2 simply collapses into D1. Our doctor’s motivation meets this condition, because he is moved by an objectively explicable reason why it would be good for him to experience the urge to get high, at least temporarily and without acting on it. So what we have constructed in modifying Frankfurt’s case is an S3-desire to experience an S1-impulse, not an S2-inclination for an S1 state.

Without some objective reason of the sort our doctor has for wanting to experience a raw urge, a mere preference or inclination for such an impulse would seem very strange indeed. But perhaps not impossible. Imagine a heterosexual male who, on experiencing what he fears might be a homosexual urge to consider other men (however slight), feels a strong

---

277Ibid, p.15.
inclination *not* to experience that ever again. Imagine that his aversion is not grounded on objective reasons (e.g. religious qualms), and it does not follow in this case from evaluating the brief homosexual urge as bad or immoral —rather, he simply finds the gay attraction revolting or undesirable, i.e. strongly prefers not to experience it, just like he happens to find peach ice cream revolting and strongly prefers not to be served it. It is difficult to keep such cases pure, without the S2-inclination becoming an evaluative desire instead, but it is just conceivable. Beyond such unlikely limiting cases, we could only find ‘perverse’ examples in which something seems to have gone wrong in the agent brain, such as someone who, for no particular reason at all, just has an inclination to experience great thirst, and so goes without water for two days. In sum, though S2 inclinations for S1 impulses are psychologically possible, they are marginal cases of complex or deviant motivation, without any immediately apparent philosophical interest.

(5) If inclinations or preferences for impulses or urges are possible, in however unusual cases, it might seem that S2 inclinations or preferences for other S2 states should be less problematic. However, there are some unexpected complications. As we saw, Frankfurt’s doctor has an S3-desire to experience an S2-state, such as a strong inclination to get cocaine by any available means (call this S2c). Now imagine instead a different doctor who has no objective clinical reason for wanting to experience this S2-state, but simply conceives a perverse inclination or a taste for S2c. It might be objected that this is tantamount to an inclination to take cocaine —perhaps not as strong as the addict’s S2c desire, but nevertheless a desire for cocaine itself. In other words, it might seem that the inclination *is* transitive in such a case, since the putatively higher-order desire *could not* be satisfied merely by experiencing S2c without also getting the cocaine, because experiencing S2c by itself would surely just be frustrating, not satisfying. Without the first doctor’s evaluative
reasons, then, it might seem that $S_2c$ itself could not become the object of any conceivable mere preference that $S_2c$ by itself would satisfy: thus an alleged preference $P$ for $S_2c$ would really be a veiled (though perhaps weaker) inclination for cocaine itself, since only acquiring $S_2c$ and taking cocaine could actually satisfy $P$.

I think we should grant that in many instances, this objection is plausible: precisely because (by hypothesis) there would no independent objective reason for desiring the $S_2$-state in a conceivable $S_2$-desire for and $S_2$-state, it is hard how the $S_2$-desire could be genuinely higher-order, i.e. how its satisfaction could be independent of the satisfaction of the $S_2$-state it supposedly desires, so that non-transitivity holds and we have two distinct orders of desire, rather than simply one complex first-order desire. But again, in highly unusual circumstances, we can imagine attitudes in which this independence is possible. For instance, suppose our doctor just conceives an inclination for $S_2c$ out of sheer curiosity. The intentional object of his higher-order inclination is a state in which cocaine is preferred for no reason except the subjective inclination for its satisfying power, but his curious attitude towards this subjective inclination is such that just experiencing the frustration of $S_2c$ without cocaine is sufficient to satisfy the curiosity. $S_2$-inclinations for $S_2$-inclinations are difficult or rare, then, because they have to be cases in which someone’s mere subjective inclination is satisfied just by the presence of another subjective preference or inclination, even if the latter is frustrated: the first-order desire without its satisfying condition (which entails frustration) must be sufficient to satisfy the higher-order desire, for them to remain distinct. Thus straightforward reflexive cases of $S_2$-desire will be highly unusual.\(^{278}\)

This analysis suggests two further complications that I can only mention here, though

---

\(^{278}\)I am deliberately waiting until Chapter IV to discuss Harry Frankfurt’s early suggestion that identification is constituted by a second-order desire (in the preference sense) that a given first-order desire should be effective or become the motive on which I actually act. This is a more complex type of second-order preference, which is not just about a first-order inclination but rather about its role in moving me to act.
they will be pursued in more depth in Chapter IV. First, it is possible to have a complex first-order desire composed of an S2-inclination that is apparently about other inclination(s): what makes it one complex desire rather than a motivation involving two distinct desires of genuinely different orders is that the former is really about the satisfaction of the latter inclination(s), and so is not independent. Consider for example a simple preference to experience, within some range of possible inclinations, only inclinations that I can easily satisfy. In this case, the general urge for pleasure or preference-satisfaction has been given a specific intentional content, which makes it a desire for easily satisfiable preferences. Clearly, this is not a desire for the preferences independent of their satisfaction, which can rest or be completed with frustrated preferences. If I live in a country where vanilla but not chocolate ice-cream is available, this desire would have me adapt my taste to vanilla instead, so that I can have a satisfied preference. This is not the same as a straightforward preference for vanilla ice-cream—it is more complex—but it is a desire for the satisfaction of whatever ice-cream desire I can satisfy. This is still only a first-order inclination, despite its complexity in referring to other inclinations.

Second, in orthodox rational decision theory, where preferences (in the minimally psychological sense of ‘revealed preferences’) must express a "complete and consistent preference ordering over all conceivable acts" (or options), it turns out that genuine preferences over preferences lead to paradoxes, or are impossible. And indeed, this result closes off a common avenue of escape that many rational decision theorists have tried to use to escape paradoxical theorems entailed by the theory in certain games (such as ‘Prisoners

---

279 The Stoics promoted a somewhat different S3-desire (based on eudaimonistic evaluations) to experience only inclinations whose satisfaction could be guaranteed (which of course meant that most of our usual inclinations were out).

Dilemma’), i.e. by introducing ‘higher-order preferences’ for cooperation. Admittedly, this
result derives from the formalisms of decision and game theory, and involves a minimal
notion of ‘preference’ that is not identical with my psychological concept of S2-motivation
(as we have seen, a limited range of unusual cases with S2-S2 states may be possible). But
arguably, the impossibility of preferences for preferences in decision theory accurately
reflects the fact that where the higher-order desire is genuinely independent, this is usually
because it is an S3-state (which cannot be represented in decision theory). In real life,
putative preferences over preferences usually turn out to be evaluative desires for
preferences, which involve a want for some preference only because of a valuable quality it
has for the agent (e.g. the fact that a certain pattern of preferences enables us to cooperate
with others for mutual benefit).

For example, an adolescent may find to his surprise that he is repulsed by the taste of
beer, and prefers soda, but he desires to acquire an inclination for beer, or to get to the point
where he would usually prefer it over soda. When we ask him why he has such a strange
desire, however, it turns out to be because his friends will think it cool to crave a cold one,
or because he sees beer-loving as part of the adult culture which he evaluates as better than
his adolescent culture, and so on. In other words, he makes a positive evaluation (however
misguided) of the taste for beer, an evaluation which on reflection he could explain to a
parent who asks why he is tasting beer when he doesn’t even like it. He sees some social
good in the desire for beer, and this good attracts him. Rather different pathological cases,
such as a patient suffering from emotional deficiency who wants to acquire some active
interests or experience strong likes and dislikes, are similar in this respect: the patient desires
(S3) such desires (S2) because he thinks it would be good for him to have them, not merely
because it would be pleasurable or preference-satisfying to have them.
(6) On the basis of the foregoing, we can easily see that there can be no stable or lasting S2-impulses or preferences for S3-evaluative desires. To be independent, such an inclination would have to be satisfiable without the satisfaction of the evaluative desire it wants to experience: but that is an attitude in conflict with the evaluative motivation at which it supposedly aims. If in having a subjective preference for an evaluative desire for some object or end X, I did not miss or feel incomplete without X, so that my preference could be satisfied by my desiring X as good without getting X itself, this attitude would prevent me from acquiring the evaluative desire. If I did somehow acquire it, then the evaluative desire would undermine the independent satisfiability of the S2-desire for it. Perhaps such a conflicted motivational state is possible, but not for long. An S2 desire for an S3 desire would either be self-stultifying or self-transforming: on being satisfied, it would become tend to become a simple S3-desire.

(7) We can largely ignore the question of S3-desires for instinctive, generalized urges or drives, after noting, pace Stoicism, that they can be rational. As Gary Watson argues, "some activities are valued only to the extent that they are objects of the appetites." For example, eating and sexual activity may be valued "only in so far as they are appetitively motivated;" this implies a judgment that "to cease to have such appetites is to lose something of worth." This means something more than that to experience certain instinctive urges and to try to satisfy them is pleasurable or preferable: it means that (at least in appropriately constrained circumstances) experiencing the urge is good in a more robust sense, as constituting part of a healthy or well-adapted human life. As Watson says, a judgment that it

---

281 Watson, "Free Agency," in Free Will, p.103.
282 Ibid, p.103.
is inherently good to have sexual appetites is necessary to explain a eunuch’s regret.\textsuperscript{283}

Similarly, a person whose psychosis prevented him from experiencing the natural urge for human companionship or friendship might feel a justifiable S3-desire for the drive he lacks. But such S3–S1 states are unusual in the sense that they are almost always responses to abnormalities in our motivational psychology, rather than a normal part of human development: they have at most a corrective value.

(8) I have already introduced the possibility of S3-desires for S2-inclinations in discussing Frankfurt’s example of the doctor and the issue of preferences over preferences. Such complex motivational states are perhaps most commonly known from experiences of evaluating some of our existing subjective preferences and inclinations as \textit{bad for us} or inimical to our long-term good, or to our chances of leading a happy life, or to our ability to fulfill our full potential. Thus the wise cigarette smoker wishes he did not feel the craving for cigarettes because he knows the health risks for him; similarly, (perhaps after counseling) an abused wife wishes that she did not prefer to stay with someone who is likely to batter her, since she knows she would be better off if she could reject him. As we will see in Chapter IV, sometimes such S3 evaluations of S2-states may express our core values and the strong evaluative contrasts involved in them. But complex motivational states of the S3–S2 variety need not express any very deep realization, existential commitment, or fundamental choice about who we are or what kind of life we want to lead: they can be as banal as a young man’s wish that he did not always seem to prefer ‘bad women’ who will toy with his emotions; as distorted as the boy’s desire to like beer; as conflicted as a racist’s wish that she did not want a date with her brother’s Chicano friend (because she judges racially mixed relations as bad); or as self-deceptive as a downtrodden person’s prayer that he should have

\textsuperscript{283}Ibid.
no ambitious desires which might make him try to get out of his fated ‘place’ in life. My point is that, although S3 states involve teleological motivation based on eudaimonistic evaluation, we cannot infer on merely on this basis alone that they express the person’s inner self, deepest motivations, or most important values and concerns. They are not necessarily evaluative motives in which, as Charles Taylor puts it, "I aspire to be a certain kind of person," or seek to define my moral identity.284

Yet as Taylor describes them, the "strong evaluations" which are supposed to explain the sense in which we can determine and are responsible for "what kind of being [one] is going to be" have the structure of S3–S2 desires. For they are motivating judgments of a "reflective kind where we evaluate our desires themselves. It is this plainly which we are tempted to think of as essential to our notion of a self."285 In such evaluations,

desires are classified in such categories as higher or lower, virtuous or vicious, more or less fulfilling, more or less refined, profound or superficial, noble or base...they are judged as belonging to qualitatively different modes of life, fragmented or integrated, alienated or free, saintly or merely human, courageous or pusillanimous, and so on....in [such] qualitative reflection there is also a use of ‘good’ or some other evaluative term for which being desired [in the S2 sense] is not sufficient; indeed, some desires or desired consummations can be judged as bad, base, ignoble, trivial, superficial, unworthy, and so on.286

So described, it is clear that the strong evaluator’s attitude towards various first-order desires on which he might act is an S3-desire, or at least its eudaimonistic judgment, which makes qualitative distinctions between ‘desirable’ and ‘undesirable’ desires on nonconsequentialist ethical grounds. The strong evaluator’s S3-desire to act on certain S2-desires and not others is thus, as we saw, "contrastive" in Taylor’s sense, meaning that conflicting options are necessarily rather than merely contingently opposed, presumably because their different

\[284\]Taylor, "Responsibility for Self," in Free Will, p.113.

\[285\]Taylor, p.112. Taylor links this phrasing explicitly to Heidegger’s formulation that Dasein is a being for whom its being is inevitably in question (p.111).

\[286\]Ibid.
intrinsic values for promoting human good are essential features of the different S2-desires which the second-order S3-desire evaluates. This is the difference made by the objectivity or normative appeal of the content at the S3 level. Merely considered at their own S2-level or in weak evaluation, however, a conflict between the same S2-desires may be merely contingent, since the impossibility of satisfying them both results only from their accidental relations to external circumstances. The contrast with David Lewis’s Humeanism is clear: for the strong evaluator, "the desirable is not defined for him by what he desires [S2], or what he desires plus a calculation of consequences; it is also defined by a qualitative characterization of desires as higher and lower, noble and base, and so on." The capacity for strong evaluation in this sense is, Taylor says, essential to any "normal human subject." But we must leave until later the analysis of whether he is correct that this class of S3–S2 desires accounts for the "depth" at which the personal self lies.

(9) As these remarks already suggest, that structure of S3-desires implies that S3-desires for other S3-desires themselves would generally be redundant. If D is an S3-desire for object X, then I could only have such an evaluative desire for D if I believed that being motivated by D would make my life fuller, more noble, more completely human, or whatever formula for flourishing we wish to use in explicating eudaimonistic judgment. And (at least in any case evident to me) I could only believe that if I also believed that X itself was necessary for my well-being in the circumstances, or desirable as good. Thus a higher-order S3 desire for D at least implies that I already have the evaluative attitude in D, if not D itself (the evaluative desire for X). But if this evaluation prompts the belief that D is

288Taylor, p.118.
289Taylor’s notion of strong evaluation will be discussed in more detail in relation to Frankfurt’s theory of identification in Chapter V.
good and thus motivates the higher-order desire for D, then in all but highly unusual cases it could motivate D itself, making the higher-order desire for it redundant. So unless something blocks an evaluative attitude from generating a desire for its object, S3–S3 desires should not arise. When they do, they may be construed as desires that first-order eudaimonistic evaluations be effective in motivating first-order desires.

In sum, then, the only really interesting cases of second-order ‘desire’ are S3–S2 desires. Although S2–S1, S2–S2, S3–S1, and S3–S3 states are also possible, they will be rare and can have value only as corrective complexes that aim to rectify some highly unusual disorder or abnormality in the agent’s motivational system. If any higher-order states that count as ‘desires’ according to the lack-model tell us anything important about the normal structure and functions of personal agency, it will be those in the S3–S2 category. In particular, it is these which must be distinguished from volitional states if we are to defend a noncompound or irreducible conception of the will as a motivational power or capacity.

II.5. Will as Conative Projection

II.5.1. Motivation vs. Intention

As we have seen, ‘desires’ or states of types S1 through S3 share a general teleological structure of ‘motivation.’ The concept of a ‘motive’ is a contested one in moral psychology, but most theorists agree that it is important to distinguish between motives and intentions. As we saw in §I.1, intention, on the dominant account today, is the first-personal understanding of what we are doing that makes our bodily movements into an action, as opposed to mere behavior. We always act under a certain description, or with a certain aspect of our behavior as our intentional object: e.g. "walking into the store to get a soda," rather than "trying to avoid someone on the sidewalk coming towards me." As this example
illustrates —and as many commentators have noted— intentions must include at least the proximate end towards which the action is consciously directed, in order to pick out the right aspect or description of the behavior which is essential to the act.

Motives, by contrast, are not particular intentions that make movements into discrete actions; rather, they typically lie behind the formation of multiple intentions, and explain why the agent fixes on the ends involved in his specific intentions to act. As Steven Sverdlik argues in a recent article, motives correspond to the beginning or major premise in a practical-syllogism reconstruction of our action, which is why we think "that motives are usually more general in content than the resulting intention," and that "a given intention (or action) could be the result of a number of different motives."\(^{290}\)

Sverdlik observes that the question of motives has rarely been addressed in "English-language philosophy" since "Action theory now largely concentrates on intention, desire, belief, and the (usually undefined) concept of motivation." However, his own analysis does little more than work out the prevailing assumption that motives are desires of some sort. Sverdlik begins with six "relatively uncontroversial" propositions about motives:

1. they are "actual psychological states or events;"
2. they are "at least part of the cause of an action or the decision to act;"
3. while they precede an action, motives "typically continue to be present or operative as the act takes place.
4. they are mentioned typically to "explain why the agent acted as she did;"
5. From the agent’s viewpoint, her motives specify "what is of value about her action" (or what valuable end it realizes).
6. "The two main types of motive seem to be emotion and desire."\(^{291}\)

He then argues that if we consider them from within the Davidsonian belief-desire model of action, "It seems clear that motives belong to the conative, desiring side of the story," since beliefs do not seem to be "even a part of the motive for an action." Factual beliefs

---


\(^{291}\) Ibid, p.335.
"concerning means, consequences, or specifications of general desires" only acquire motivating force when they are connected with some desire. Of course, anti-Humean theorists have defended the idea that some evaluative beliefs do constitute motives to act, but Sverdlik insists that internalists about moral obligation can hold that the motive is "some separate state of desire which is entailed by the belief" about one’s obligations.292 This is roughly to say, in my terms, that internalists do not need anything more than a kind of S3-state in which the evaluation is sufficient for the desire. Similarly, "the emotions we regard as motives characteristically give rise to desires to act in ways related to the content of the emotions,"293 e.g. fear gives rise to the desire to flee.

If this is right, then beliefs and emotions become motivating only by way of desires, and the best hypothesis is that the motive is the "ultimate desire" from which the end in our intention is derived.294 And since "the identity of an action rests partly on what intention it incorporates," on all the leading views about act individuation, "all intentional actions (or all actions under the descriptions in which they are intentional) have a motive."295 In sum, on Sverdlik’s view, which is representative of the current literature, beliefs and emotions cause desires, and these in turn motivate intentions, which get their ends from at least one (though possibly more than one) ultimate desire(s): "intentions are formed as a way of pursuing the agent’s goals or desires."296 Intentions are thus instrumentally explicable relative to certain terminal desires.

293Ibid, p.336. The example Sverdlik gives is envy, which "gives rise to the desire to remove the envied object from another person." He acknowledges that some emotions, such as grief, may not give rise to desires, but says that then they fail to constitute motives.
295Ibid, p.337.
In my view, this sort of analysis which reduces all motivation to desire—even including S3 desires—ignores or excludes a fundamentally different kind of motivation, which does not have the teleological form or ‘lack-structure’ that S1-S3 states have in common. Indeed, the teleological model stemming from the Greeks has so dominated all Western moral psychology that the presence and significance of this other kind of motivation has been systematically obscured, leaving it as virtually uncharted territory. As we will see, although Kant’s account of moral motivation enters this territory, even he does not recognize its breadth nor its general structure, because he is focused on construing it simply as a motivation to act that has a purely formal ground. Thus neo-Kantians today concentrate on arguing that moral judgments and beliefs can motivate without depending on pre-existing desires in the agent’s ‘internal set,’ while their Humean opponents argue that all motivation must derive originally from S1-S2 desires. In the process, neither recognize that there is a distinctive kind of motivation which is neither derivative from desires, nor simply constituted by moral beliefs (nor even by desires following contingently from formal judgments, as Sverdlik suggests).

For the claim that there is a non-desiderative kind of motivation to be significant, of course, it has to be made relative to a substantive account of desires, like the one I have given in §1-3 above. This is not to deny, as Pink says, that "There surely is...a sense of ‘desire’ that applies whenever we are motivated to act." for we could define ‘desire’ in an empty, merely formal fashion as ‘any state in which an agent is motivated to pursue some goal or end.’ Similarly, we could conceive ‘teleology’ in a weak or formal sense merely as any psychological state in which an agent "conceiv[es] of the future as including a state of affairs which is an end to be produced, where this end is provided in the propositional

---

content of a belief or desire" [also in the weak sense]. But in these minimal or empty formal senses, it is vacuously true that all motivation is ‘teleological’ or ‘desiderative,’ and hence these senses are too weak to allow expression of substantive theses about motivation like Hume’s or Aristotle’s. To express their views, we need strong or substantive senses of desire: e.g. Hume holds that all motivation derives from S2 desires, and Aristotle holds that motivation derives primarily S3-type desires. The Platonic ‘lack-model’ I described earlier similarly gives a substantive sense to the notion of teleology as a structure of motivation, or a way of having ends, rather than just any type of goal-directedness.

This distinction between formal and substantive senses of ‘desire’ is also made by Thomas Nagel in his early argument against Williams’s Humean model of motivation:

The claim that a desire underlies every act is true only if desires are taken to include motivated as well as unmotivated desires, and it is true only in the sense that whatever may be the motivation for someone’s intentional pursuit of a goal, it becomes in virtue of his pursuit ipso facto appropriate to ascribe to him a desire for that goal. But if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit, and it is by no means obvious that a desire must enter into this further explanation. Although it will no doubt be generally admitted that some desires are motivated, the issue is whether another desire always lies behind the motivated one, or whether sometimes the motivation of the initial desire involves no reference to another, unmotivated desire.

‘Desire’ is either motivated and unmotivated, and it is trivial or vacuous to say that all motivation must involve ‘desire’ in an empty sense neutral between these two qualifications. Motivated desire is a determined rather than determining variable, so its presence may be due to entirely non-desiderative prior motivations. The Humean needs to claim that all motivation is ultimately rooted in unmotivated desires (of the S2 types, which is non-evaluative or ‘emotivistic’ and makes no objective value-claims), and this is a much less plausible or obvious thesis. For even S3 desires would count as ‘motivated’ in Nagel’s

---

sense, for in these cases, the desire for some end E is the result of certain evaluative reasons, and hence the desire itself "obviously cannot be among the conditions for the presence of those reasons." Similarly, Nagel suggests that the structure of practical reasons may explain "those general desires which embody our acceptance or the principles of practical reason." Thus postulating such general desires will not help a defender of rational egoism answer altruist objections, because if ‘rational egoism’ is construed so that it can always be supplemented by adding a subjective preference for any type of virtuous end or moral restraint on action, it becomes proof against all possible counterexamples only at the price of ceasing to embody any substantive theory, or becoming wholly vacuous.

Consider one further example of this point in a very different context. In the midst of explaining one of Derrida’s basic theses, John Caputo makes the Nietzschian assertion that there is no motivation not originally driven by the agent’s need for satisfaction, there are only more or less "selfish" forms of it:

> We are all more or less narcissistic, for that is what the agent/subject is. The agent, Aristotle and the medievals said, acts for its own good. If the agent expends all its energies on the other without return, that is after all what the agent wants, and that is how the agent gets her kicks.

The fallacy here is an equivocation on "wanting." It would only be true a priori that anything that an agent works for is "what the agent wants" if "wanting X" is emptily defined

---

300 Ibid, p.30. Nagel adds, interestingly, that "the temptation to postulate a desire at the root of every motivation is similar to the temptation to postulate a belief behind every inference. Now we can see that the reply in both cases is the same: that this is true in the trivial sense that a desire or belief is always present when reasons motivate or convince—but not that the desire or belief explain the motivation or conclusion, or provides a reason for it.” Nagel gives as an example an inference in accordance with modes ponens. Although the inferrer believes in this principle of logic, "the belief that this principle is true is certainly not among the conditions for having reasons to draw conclusions in accordance with it.” Rather, the belief in it follows from those conditions and reasons just as does the inference in accordance with it (pp. 30-31).

301 Ibid, p.31.

302 Deconstruction in a Nutshell: A Conversation with Jacques Derrida, ed. John D. Caputo (Fordham University Press, 1997), p.148. Note that by "the medievals," Caputo must mean Aquinas, for certainly Scotus and several others would protest most vehemently against the Nietzschian position he is derives from Derrida and tries to whitewash here (see Chap. III §3).
just as "seeking to realize X" or "being motivated to pursue X." But such a definition is so non-substantive that it entirely fails to capture the view Caputo means to defend, since under this definition, it is conceivable that an agent could "want" an end in an absolutely "non-narcissistic" fashion, which is not what Caputo means. The substantive issue is how the motivation arises, not whether, once the agent has an end, she derives satisfaction from realizing it. So Caputo has no argument here, and as the rest of the chapter will show, his view is probably wrong.

II.5.2. Projection as a Different Kind of Motivation

The structure of the alternate kind of motivation that falls outside the dominant belief-desire paradigm can thus be made clear only by contrast with the ‘lack structure’ of motivation which we have identified as the distinctive mark of ‘desire’ or teleological motivation in its substantive, classical sense. Relative to this, the claim that there are instances of non-teleological motivation means that a human agent can be motivated to pursue ends because she has given herself these ends or adopted these goals as her own, rather than because the agent is ‘drawn’ by their pre-existing goodness or apprehended value for the her well-being. In this largely unrecognized motivational structure, which I call ‘projective,’303 the motivation first arises not from the intentional object but from the agent: the agent projects himself towards the object or state of affairs that will be the goal of his actions, rather than simply needing or wanting this state or end in the sense of requiring something of its kind to requite a lack or fill a felt absence in his well-being. As Peter Bertocci argues, "the quality of purposive activity that a person undergoes as a feeling-

303The term is Heideggerian, but at this point, I use ‘projection’ because it connotes throwing something out ahead of ourselves, or putting something in the functional position of being our goal or end, instead of simply finding it in that position because of its inherent qualities.
wanting agent must be differentiated from the quality of purposeful experience whereby he organizes his feelings and wants in accordance with some consciously held goal."\(^{304}\) We should add that for agency to be ‘purposeful’ in this sense, its goal cannot not fixed as the agent’s goal or practical object without her participation. Her experience is not simply a more complex form of being purposively drawn towards an end, because the object only becomes her goal by her through her making it an end for her activity.

This is not yet to say that such projective motivation is ‘freely created’ in a libertarian sense, and certainly not to say that it arises from no grounds at all. As Bertocci suggests, "will-agency does not occur within a vacuum...[since] the person at choice-point confronts habits, attitudes, traits within his acquired nature,"\(^{305}\) which influence his projection of goals. Moreover, projection characteristically involves rational considerations. Like S3-desires, projective motivation usually has ‘reasons’ in the sense of cognitive grounds with objective purport, but unlike the evaluations in S3-desires, which evaluate the object or goal of possible action as ‘good’ in the sense of contributing to our well-being or proper functioning, broadly understood, the judgments underlying ‘projections’ evaluate the intrinsic worthiness or value of projecting this or that as an end itself —an evaluation that does not depend (solely) on whether this end satisfies a want or contributes towards the completion of the agent’s well-being. Thus projection is a conative or motivational state of tendency towards an end, which is nevertheless not teleological or ‘broadly desiderative’ in any substantive sense. For as Pink argues, "Desire formation, characteristically, is passive passion, not active action,"\(^{306}\) whereas projection is enacted, actively formed motivation for

\(^{304}\)Peter A. Bertocci, *The Person and Primary Emotions* (Springer-Verlag, 1988), p.5.

\(^{305}\)Ibid, p.7.

\(^{306}\)Pink, *The Psychology of Freedom*, p.19. Pink also notes that whereas "intention is always directed at agency," "I can want E without wanting E to be brought about by my agency" (p.18). So desires are not as (continued...
closely related to agency as our intentions, which are both formed by and aim at our actions.\textsuperscript{307}

To start with a deceptively simple example, I have motivated myself to work out a philosophical problem through committing myself to this task, or projecting its solution as my end. In this case, my motivation does not follow from a prior attraction of this end, but the reverse: the motivation precedes and determines its becoming my end. I do not believe that my well-being will be less without the solution to this question—or if I do believe this for other reasons, and so conceive an S3-desire for the solution, this desire is independent of my projecting the solution as my aim, and so we have a case of overdetermination. In most other respects, the motivation formed by projection functions like a desire: for example, it leads me to form various specific intentions, such as to read an article on the topic in order that I might understand it better. But the motivation does not have the ‘lack-structure’ of a desire: to put it in dynamic metaphors, the motive originates from the agent ‘pushing’ the goal out in front of him, rather than by the beguiling end ‘pulling’ the agent towards it.

This distinction between desires and projective motives may be framed in terms of familiar distinctions in ‘direction of fit.’ Contemporary action theorists typically distinguish between cognitive and conative intentional states by referring to differences in their direction of ‘fit’ with the world, or extramental reality. Belief-contents are said to be guided by the world, or to fit themselves to reality as we understand it, while desires seek to bring about states of affairs that would agree with the content desired (or seek to fit the world to that which is desired). Note that if we accept the completeness of this model, which underpins

\textsuperscript{306}(...continued)
closely related to agency as our intentions, which are both formed by and aim at our actions.

\textsuperscript{307}This might seem at odds with Thomas Nagel’s argument that there are both "motivated and unmotivated desires," or appetites that "simply assail us" as well as desires motivated by other more basic desires and possibly other considerations (\textit{The Possibility of Altruism}, p.29). But in response, we could simply say that the motivation produced in projection is a \textit{motivated desire} that is not formed on the basis of any unmotivated desires, and this would be perfectly compatible with Nagel’s model.

145
the dominant instrumentalist belief-desire accounts of motivation, then it would appear that
cognitive and conative functions are mutually exclusive, since no single state could involve
cognitive judgments and conative ‘movement’ without having contradictory directions of fit.
If this were right, then volitional states in the indivisible (or non-composite) ‘middle-soul’
sense described earlier would be ruled out.

But this belief-desire analysis is too simplistic: it only considers the direction of fit
between the intentional content and the world, and fails to consider the relation between the
agent of the intending and the content intended. Even though philosophers of mind typically
recognize a distinction between the state of believing or desiring, and the content believed or
desired, this distinction has played little role in contemporary action theory. Yet there is a
direction-of-fit distinction to be drawn at the agent-content level, as well as at the content-
world level. As we have seen, classical philosophers such as Plato and Aristotle recognized
the agent’s being ‘drawn’ by (or fit to) content (rather than the anticipation of fitting the
actual world to the content) as the distinctive mark of desire. The ‘lack-structure’ analysis of
desire just means that desires have what I called a teleological agent-content fit: it is
essential to desires of all kinds (appetites, passions, or rational desires) that the agent’s
motivation is fit to the content desired (either by brute ‘pull,’ or by being guided by
subjective or objective evaluations of its object). Thus in general, where the arrows indicate
direction of fit, desires have this schema:

 Desire:  Agent → Content ← World

Despite their difference in content-world fit, beliefs are like desires in their agent-content
fit: thus beliefs do not appear to be under our direct voluntary control. Plato and Aristotle
exploited this similarity to emphasize the links between desire and beliefs or judgments
about the human good. As with original (or non-derivative) desiring, the agent’s believing some content seems to be passively directed by belief-forming faculties that consider different contents eligible for belief in light of all sorts of evidence. Thus beliefs or judgments have the following schema:

**Belief:** Agent → Content → World

By contrast, projection is unlike both belief and desire in its agent-content direction of fit: in projection, the intended content is ‘fit to’ or guided by the agent’s active intending of it, rather than the reverse. When the resulting state has a conative content-world fit, it will constitute a motive and guide the formation of particular intentions to act: it is a state of projective motivation (or ‘project’ for short). It may also be possible to project intentional contents that have cognitive content-world fits: this provides one interesting way of interpreting states of faith and their difference from ordinary beliefs (as well as explaining some of the intuitive resistance to accepting the involuntariness of all ‘beliefs’).

Thus we have the following schema:

**Projects:** Agent ← Content ← World

**Faith:** Agent ← Content ← World

Projects are active in the same way as states of faith so conceived: they differ only in their content-world fit, since states of faith are appropriately about contents that make reference to reality as it is, and that we cannot usually will or give ourselves as intention-guiding ends for

---

308I can in principle remain neutral about this way of construing faith. But it is useful to complete the scheme of four permutations made possible by adding the two agent-content directions of fit, and faith so construed provides what I think is an instructive analog to cases of projective motivation.
On this topic, see John Hick’s work on the cognitive significance or factual claim-making functions of religious belief, and in particular his work on “eschatological verification.”

I might also be motivated to pursue it, for example, by the desire for employment, but this would be an additional motive, rather than an exclusive alternative. Projective and desiderative motivation often overdetermine pursuit of a single unified end, though their relation remains contingent and can come apart.

In this respect, states of faith have a cognitive function while projects have a conative one. But both are actively undertaken by the agent, and are cognitive in the additional sense that they can reflect her reasons or grounds for such an undertaking.

This analysis helps make clearer both how projective motivation can be conative without being desiderative, and how it can be made to fit (or be guided by) considerations the agent has other than those which typically fit his motivation to a presented content, as in desire. In such cases—which I label S4 states, in anticipation of distinguishing them from emotions (S5) and evaluative attitudes (S6)—the agent’s approach to the goal is not explained in terms of prior attraction towards it because of its perceived power to satisfy preferences or its inherent value for her well-being, or its instrumental value for either of these. Rather, the agent’s motivation to pursue the goal (which may still be weaker or stronger) is explained by his determination to take it as her end, and from the reasons which informed this projection. For projection never occurs ex nihilo. For example, I have rational grounds for determining myself to solve my question, because the question is inherently important in my field, because a commitment to pursuing it is noble and potentially enlightening even if the question is not adequately resolved, and so on. Similarly, the needs of other persons and their views regarding possible projects can give me grounds for projection. But these reasons are not the kind that create a sense that the solution promises to fulfill an inadequacy in my satisfaction or well-being; they are not considerations that foster a desire, even of the S3-kind, for the solution. Such cases allow us to isolate projection from other main types of motivation, namely S1-S3 states.

---

309 On this topic, see John Hick’s work on the cognitive significance or factual claim-making functions of religious belief, and in particular his work on "eschatological verification."

310 I might also be motivated to pursue it, for example, by the desire for employment, but this would be an additional motive, rather than an exclusive alternative. Projective and desiderative motivation often overdetermine pursuit of a single unified end, though their relation remains contingent and can come apart.
Certain sorts of radically evil and perverse ends that have long posed problems for teleological accounts of motivation become easier to explain (though admittedly not without some remaining mystery) if we accept the existence of projective motivation. For example, if spite is a motivational state in which the spiteful agent is bent on harming another person to any degree possible, even if he can only achieve a little harm at great cost to everything else that matters to him in life, no non-self-deceived judgment about his own happiness can produce such motivation as an S3-desire. Of course the agent can draw satisfaction from inflicting harm on his target, but only because he is antecedently committed to doing it whatever damage he can, no matter at what price: thus the anticipated satisfaction cannot itself be the root of this commitment. His motivation is pure in a way similar to that of someone demonstrating perfect neighbor-love: he is prepared to sacrifice his own well-being absolutely for the sake of his goal. It is difficult to see how such a goal could become one’s end except projectively. Naturally his mission against the other person cannot be for no reason if it is not utterly insane: it must have some grounds, however inadequate they might seem to third parties. But by themselves they do not control the agent’s resolve: e.g. (to take an example from the movie Cape Fear), to abhor my lawyer for betraying me is not in itself to live solely to discomfort him, to work for nothing but his destruction.

There are (probably) even cases where dissatisfaction itself is the agent’s primary end, or where lack and deficiency rather than their fulfillment is the goal: in such cases, virtually by definition the motivation cannot be desiderative. Consider Sartre’s description of a person with an "inferiority complex," who takes his being-for-others as a basis for self-abhorrence:

the inferiority that is felt and lived is the chosen instrument to make us comparable to a thing....But it is evident that it must be lived in accordance with the nature which we confer on it by this choice—i.e. in shame, anger, and bitterness. Thus to choose inferiority does not mean to be sweetly contented with an aurea mediocritas; it is to produce and assume the rebellion and despair which constitute the revelation of this inferiority. For example, I can persist in manifesting myself in a certain kind of
employment because I am inferior in it... It is this fruitless effort which I have chosen, simply because it is fruitless—either because I prefer to be the last rather than to be lost in the mass or because I have chosen discouragement and shame as the best means of attaining being.

If it really is possible for an agent to be so motivated, then she clearly cannot desire her inferiority in the sense of taking satisfaction in it. We cannot have a second-order desire to experience unsatisfied desires for approval and success, because as we saw, for a desire to be truly second-order, its satisfaction must be independent of the satisfaction or non-satisfaction of the first-order desires to which it refers. Sartre’s agent does not want any satisfaction in being regarded as inferior: ironically, her aim would be frustrated if she were happy in attaining it. Sartre imagines an existential ground for such a project that is just conceivable: the agent is desperate enough to see inferiority as a viable way to be assured of some definite meaning in her life, no matter what the cost in happiness. We might suggest more plausible grounds for this project instead, such as years of neglect and abuse by others, or failure in ambitions earnestly pursued for success, but the whatever its ground, such motivation cannot be teleological.

Thus although it involves giving ourselves ends instead of having them assigned to us in desire, projective motivation is never simply generated ‘at will’ without grounds. We do not just give ourselves an end and acquire a motivation to pursue it without any antecedent conditions. There must be grounds, yet these grounds do not necessitate motivation and do not themselves constitute motivation without projection. Ordinarily, the cognitive and projective sides may form a whole in our experience, and seem inseparable, although they can be distinguished in philosophical abstraction.

But if this is so, can we really distinguish projection from the kind of S3-desire that flows immediately from eudaimonistic judgments? For as scholars of ancient ethics remind us, the type of "happiness" we are assumed to desire on this model is of course not "a determinate
and specific state, [or] a state of feeling good about something." It refers to a general and indeterminate sort of ‘good’ for the agent that can go beyond satisfaction in any narrow sense to include her living a well-integrated complete life, a "concern for her life being as it should." Could not any grounds I have for "projecting" an end be understood as referring to that final good, and the motivation then be at least formally regarded as a desire for that general telos? This is suggested, for example, by Julia Annas’s remark that "For the ancients, desire, orexis, is the most general kind of motivation to do something that we can have. It covers wanting of various kinds, and also covers motivation generated by reasons, including ethical reasons."

Yet, as I have argued, if the claim that motivation is desiderative is to have any content, then desire or orexis cannot be used simply as a shorthand for ‘being motivated to pursue some end.’ And as we saw, in the De Anima at least, Aristotle follows Plato in giving a substantive analysis of orexis as being drawn towards an end that is qualified as fulfilling some lack in the agent. Only in this context can the claim made in ancient ethics that "all action is aimed at some good" for the agent be more than a tautology, and thus able to do some work in ethical theory. Of course ethical dispositions and moderate attitudes that are essential to living a fully human life may be precisely what is perceived as lacking, or not fully realized, and thus desired in S3-fashion. But the underlying sense that the good is a kind of equilibrium and that I am drawn towards my end because it will help restore that equilibrium is necessary to understand what it means, as Annas puts it, that "I do all these things [whatever I do], simple and complex, because I see them as contributing to my telos, 

---

312 Ibid, p.28.
313 Ibid, p.35.
my final end which is my final good."\textsuperscript{315} In other words, in ancient ethics, the lack-analysis provides the sense in which desiderative motivation refers to my good; this implies that we cannot just treat the ‘agent’s good’—the end she can desire—as an unbound variable whose value can be filled in by any sort of end whatsoever. For this reason, I cannot agree with Annas’s claim that, given the generality of orexis, the doctrine of a final good that is the end for all our motivations "does not imply that there is any special kind of motivation that we have to our ends."\textsuperscript{316}

In particular, it may turn out that the ends typically pursued by an agent with various virtues may be hard to fit into the lack-structure of desire: if so, this will imply that when Aristotle and other ancient ethicists introduce virtues, whether they recognize it or not, they are introducing motivational phenomena that cannot be accommodated within orexis understood in the substantive sense provided by the De Anima.\textsuperscript{317} The next section takes up this issue.

\section{II.5.3. Virtue, Goods Internal to Practices, and Projection}

\textit{—Aristotle on Virtue and the Paradox of Eudaimonism}

At this point, neo-Aristotelians are bound to raise a number of objections. For in sketching the contrast between S3-desire and projection, and equating eudaimonistic

\textsuperscript{315}Ibid, p.35.
\textsuperscript{316}Ibid, p.35.
\textsuperscript{317}For this reason, Annas is also too quick in saying that "If our final end is in fact virtue, then virtue is our ultimate object of desire. This does not reduce virtuous motivation to wanting; rather it brings under the umbrella of desire the way we are motivated by reasons of virtue" (p.35). This claim may well reflect Aristotle’s own conviction, but it ignores that there is a substantive issue as to whether the motives involved in virtue can in fact be brought under this desiderative umbrella, as Annas (and probably Aristotle) simply assume.
motivation with the former, I may seem to be implying that virtue as classically conceived bottoms out in the agent’s concern for her own well-being, which, however enlightened, is still a basically egoistic concern. But this is a long-exposed fallacy resulting from modern prejudices and insufficiently careful reading of the ancients. For Aristotle says that we may do what, outwardly, the virtuous person would do without doing it from virtuous motives:

...if acts that are done in accordance with the excellences [aretai] have themselves a certain character it does not follow that they are done justly or temperately. The agent must also be in a certain condition when he does them; in the first place he must have knowledge, secondly he must choose the acts and choose them for their own sakes, and thirdly his action must proceed from a firm and unchangeable character.

Thus a virtuous motivation is a stable disposition to choose the right act for the circumstances because it is noble — not because it makes the agent better-off. For example, the brave man who encounters terrible dangers "will face them for the sake of what is noble, for this is the end of excellence [virtue]." The virtuous person values a certain end (such as courage) intrinsically, not instrumentally, and only this ‘pure’ motivation contributes to eudaimonia. So it seems that my earlier analysis of Aristotle was unfair in concluding that

---

318Note that, as Annas herself argues, almost all ancient ethical doctrines are eudaimonistic in form.

319By this Aristotle is apparently referring to the Platonic idea that a person with a virtue must know the good aimed at by that virtue, since he goes on to downplay the significance of this condition relative to those produced by practicing virtuous acts rather than simply reflecting on the good. But Aristotle may also be referring to the point made later in his discussion of voluntariness (Ethics III.1) that to be held accountable, the agent must know the circumstances of his act sufficiently for him to intend it as this particular act under one description, rather than another description of his behavior: we are obviously not virtuous if the description under which our behavior would be a virtuous act is not the one we intended in performing it.

320Aristotle, Nicomachean Ethics, II.4, 1105a.29-33, p.1746.


322Aristotle may have this point in mind when he argues that "every excellence both brings into a good condition the thing of which it is the excellence and makes the work of that thing be done well; e.g. the excellence of the eye makes both the eye and its work good, for it is by the excellence of the eye that we see well. Similarly, the excellence of the horse makes a horse both good in itself and good at running...the excellence of man also will be the state which makes a man good and which makes him do his work well" (Ibid, II.6, 1106a.15-23, p.1747). A horse does not perform well just in order to qualify as a good horse; when it needs constant incentive, it is not yet a good horse. But when it performs well for no other end than performing well, it is good. Similarly, human persons are properly functioning as thus "happy" in the fullest sense only (continued...)
orexis includes only the type of states I have described as S1, S2, and S3 desires. Instead, his Ethics implies that other types of desire are possible which, unlike S3-desires, do not even ultimately make everything else merely instrumental relative to the agent’s own flourishing or well-being.

This is Annas’s main point in arguing that ancient ethics is not egoistic in starting from reflection on one’s own life and what is needed to improve it:

For what I have to develop, in order successfully to achieve my final good, are the virtues...Some have a direct connection with the good of others, most prominently justice, which may involve surrendering goods I want to others because they have a just claim on them; but all the virtues are dispositions do the right thing [for its own sake], where this is established in ways that are independent of my own interests.\(^\text{323}\)

So put, however, this is bound to remind us of familiar paradoxes in rational decision theory surrounding cases (such as deterrence) in which it seems to be ultimately in someone’s interest to be willing not to act from self-interested motives. These paradoxes are studied precisely because they seem (from the game-theoretic perspective) so closely related to the function of moral restrictions in preventing collective systems of self-interested behavior from becoming entirely self-defeating. Yet the fatal flaw of this approach is always that these observations could make altruistic motivation (for example) a rational strategy from the self-interested perspective only if the individual could somehow trick or deceive herself into genuinely giving up selfishness as her maxim in order to gain from this a benefit that, once she is in the state necessary to reap it, she would no longer desire as her final end, or

---

322(...continued)
when they are purely motivated to act for the sake of the noble.

323Annas, The Morality of Happiness, "The Good of Others," p.223. I should note here my view that this book is a monumental achievement in scholarship on ancient ethics in general. I chose Annas as my primary interlocutor here only because she makes the best attempt I know to defend eudaimonistic ethics from the charge of egoism. In taking issue with this one aspect of Annas’s multifaceted work, I do not mean to imply disagreement with her convincing main thesis that ancient ethics concerns largely the same things that moderns mean by morality, and that despite major structural differences in approach, the gap is not as large as both sympathizers with ancient ethics (like Williams and MacIntyre) and its legion of critics have made it seem.
that would no longer have the same value for her.

Annas obviously (and rightly) is not appealing to this sort of self-deception to explain how virtues fit into a eudaimonistic moral psychology. Instead her point seems to be simply that something can form part of my good without my intending it for this reason: "There is no reason, prima facie, why the good of others cannot matter to me independently of my own interests, just because it is introduced as something required by my final good."

Similarly, in response to remaining doubts about egoism, she writes: "But it is no part of the theories we have seen so far that its forming part of my good is the reason why I should care about the good of others." Thus "[t]he straightforward claim that an ethics of virtue is egoistic, since the agent is concerned about developing her virtues as a way of achieving her final end, is straightforwardly mistaken....An ethics of virtue is therefore at most formally self-centered or egoistic; its content can be fully as other-regarding as that of other systems of ethics."

This explanation seems plausible at first, but it has a major problem. If it is right, what then will be the point, for me, of recognizing, as Annas puts it, that "Achieving my final good, happiness, or whatever that turns out to be, will involve respecting and perhaps furthering the good of others"? On her account, it would seem that this discovery made while reflecting on my life could only be of speculative interest for me, because if it becomes (part of) my motive for respecting others, I will precisely have failed to respect them as valuable for their own sake, thus at the same time stymied the improvement I hoped to achieve in my own life (the self-defeat paradox of egoism, once again). Unless via the

---

324Ibid, p.223.
325Ibid, p.224.
327Ibid, p.223.
kind of self-manipulation implied by the game-theoretic approach, then how could recognition of the eudaimonistic value of "pure" or non-egoistic motivation do any practical work from the agent’s perspective?

The problem here can be framed more rigorously as follows. It is possible for the following propositions, for instance, to be true in conjunction:

(1) Alfred cares about Jane for her own sake, or as an end-in-herself, and
(2) This caring makes Alfred’s life better, happier, or more fulfilled.

But the relevant question is whether the truth of (2) can somehow figure in Alfred’s motive for caring about Jane in (1). In other words, can his presumably ultimate motivational orientation towards his own eudaimonia or flourishing motivate, lead to, or explain his caring about Jane in a way that makes him prepared to sacrifice other good things of his, such as external pleasures, power, or even perhaps life itself, if it came to it, for Jane’s sake?

It would be a mistake to say that this is impossible on the assumption that it is simply contradictory for concern about any terminal end or intrinsic value to be related motivationally to eudaimonia. For the issue is not whether Alfred can undertake the care described in (1) as a means to his own happiness or flourishing — that obviously would be contradictory. Rather, the relevant issue is whether he can come to care about Jane as an end having intrinsic or terminal value and yet do this because he recognizes that so caring will be a constitutive part of his happiness or flourishing. This proposal seems to be the best (and perhaps the only) possible way of reconciling Aristotle’s orexis-theory of motivation in the De Anima and the claims he makes in his Ethics about the motivational structure of the virtues, which contemporary commentators such as Annas have so strongly emphasized. The question, then, is whether desire for something other than my own eudaimonia (such as Jane’s well-being) can be motivated by recognition that such a desire (at least when stable)
will be part of my eudaimonia.

Indeed it seems possible that I can come to desire some end, for its own sake and not as a means to anything else, and yet do so because I realize or judge that this end itself will be an integral part of a good life (in the sense of a happy, fulfilling life that exemplifies the proper use or development of my full human potentialities). The belief that achieving a given end will contribute to my eudaimonia may trigger my desire for this end without that implying that the end is valued only instrumentally: on the contrary, I judge that realizing this end will yield well-being, or make my life better (or however else we cash out eudaimonia), and no other function it might serve is needed to make it desirable to me (even if I recognize that it has other effects that may be good in other senses). In my terms, this just amounts to recognizing that we can have an S3-desire for an end that has intrinsic value for us: we desire it because we see it as a constitutive part of our eudaimonia. Thus an S3-desire for some end X does not necessarily imply that we only value X instrumentally, as a means to the distinct, separate end of our own flourishing.\textsuperscript{328} On the contrary, S3-desire is one of the basic ways in which we acquire goals that are desired for their own sake, or pursued as ends-in-themselves. So there is nothing inherently impossible about the following form:

\begin{quote}
\begin{enumerate}
\item Jane values something X intrinsically, or pursues it for its own sake, and
\item X became an end-in-itself for Jane because she judged it to be a constitutive part of her overall well-being in the sense of flourishing over a complete life.
\end{enumerate}
\end{quote}

There are many instances of X where this conjunction could plausibly apply. And furthermore, we must also grant that they include a number of cases where it is some activity itself that is substituted for X (e.g. where X = running a mile in under five minutes, or studying galactic astronomy, or tutoring an immigrant in English).

\textsuperscript{328} This does not mean, however, that some S3-desires are in no sense "egoistic." For valuing something instrumentally or only as a means to one’s own interests is not the only way to value something egoistically.
But the paradox has not been dissolved: we have simply not yet pinned down the exact point where it threatens Aristotle’s account (and other eudaimonistic ancient ethical theories). The problem for these theories arises because a desire D for an end E cannot itself be construed as an "activity" for which we can have a second-order S3-desire because we recognize that D and the pursuit of E it motivates will contribute to eudaimonia. For the reason we can find plausible instances of the form given by (3)-(4), including some instances that are themselves activities, is that in these instances, it seems plausible that the *end itself, as opposed to the pursuit of it or the desire for it,*\(^\text{329}\) could contribute to the agent’s flourishing or overall fulfillment in life. Where it is the other way around, and the motivation to pursue an end rather than the end itself contributes to the agent’s happiness, this cannot explain the motivation and pursuit. There are obviously many instances in which:

(A) It does not seem so plausible that realizing the end itself (whether it is an activity or not) can add much to the agent’s well-being or eudaimonia (at least directly), since that end mainly concerns the state of other persons, or other things, or states of the agent not vitally relevant to her own well-being.

(B) But being motivated to realize that end and pursuing it could be a constitutive part of the agent’s eudaimonia. This will include both purely altruistic ends, such as the well-being of a stranger, and many others, such as ends acquired through friendship and associations, and the ends pursued by some practices, such as fixing someone’s teeth, or building new houses for others, or giving a guilty client a successful defense.\(^\text{330}\)

---

\(^\text{329}\)Note that in the case of activity-ends, this distinction still applies. We can still aspire to an activity, and desire to perform an activity, without yet being able to perform this activity, or to engage in it fully or in the right way. Of course, if we define desiring some end itself as a kind of activity, than in the case of this special sort of activity, the pursuit/achievement distinction tends to collapse (although even here, as we have seen in analyzing second-order desires, it is possible to want to have some desire, or to start some pursuit, without having achieved this oretic state). But an ‘activity’ of this oretic sort could not have itself have intrinsic value for a person’s eudaimonia on Aristotle’s theory, because as we have seen, a desire always signals a lack or gap in one’s well-being, namely the need for precisely the contribution that the desired end will fill.

\(^\text{330}\)Of course, no one could engage in any practice seriously without some pride in its intended outcome, and this pride, whether it is a moral or aesthetic satisfaction, may always add something to the agent’s well-being. But as we will see, it is not these contributions that the products of the practice may make that generally (continued...)
In such cases, as many defenders of Aristotle have not seen clearly, the (3)-(4) form cannot explain how we acquired the motivation to pursue this end. For these instances, the Aristotelian would clearly need something like the following form:

(5) Jane cares about Alfred’s well-being as an end-in-itself.

(6) Pursuing this end (as considered independently of the realization of the end itself) contributes directly to, or is a part of, Jane’s eudaimonia.

Now the conjunction of (5) & (6) can certainly be true, as (A) and (B) suggest, and the same holds if we substitute any of many other for ends we might put in place of Alfred’s flourishing in (5) and (6). But this (5)-(6) form, unlike the (3)-(4) form, does not allow for a eudaimonistic explanation of how Jane acquired this end, or became motivated to pursue it. Granted that pursuing it is good for her, this cannot be not what motivates Jane, since by hypothesis it is Alfred’s doing well that she takes as her terminal goal or pursues for its own sake. Yet her coming to value this end intrinsically cannot be a result of desire, since the realization of this end itself, as abstracted from the pursuit of it, is not any part of her eudaimonia, and thus she cannot apprehend its being unrealized as a lack before being committed to the end. Once she acquires this end, pursuit of it can certainly have eudaimonistic consequences for her, but this cannot have had anything to do with why she came to pursue this end in the first place.

Here, then, is where the teeth in the paradox start to bite. It would be paradoxical to say that Jane recognizes Alfred’s well-being as a terminal end for her, and does so because she realizes that being moved by a desire for this end will be good for her or improve her life. For this would mean that in fact Jane saw Alfred just as an occasion to develop in herself a character-trait that is part of a good life, rather than seeing him as an end-in-himself: and so

330(...continued)

concern the virtue theorist in her argument that the having and exercising the virtues contributes to eudaimonia.
of course, as the virtue theorist herself is always first to point out, Jane would then in fact fail to acquire precisely the virtue she would be aiming at, i.e. the motivational state that actually contributes to human flourishing. To avoid this paradox, we can only link Jane’s care for Alfred to her eudaimonia in the way that form (5)-(6) proposes. But this form is then incompatible with a eudaimonistic account of how Jane acquired her goal, or came to pursue Alfred’s good as a terminal end. This is a serious dilemma for the Aristotelian: (a) if she appeals to its link between Jane’s other-regarding motive and her eudaimonia to explain her motivation, then she runs into contradiction, but (b) otherwise, the origin of the motivation that in fact contributes to Jane’s eudaimonia can only be explained in a non-teleological, anti-orexic fashion. Only by introducing projective motivation can we resolve the paradox.

Other analyses unwittingly support this conclusion. In an insightful essay, Richard Kraut argues that for Aristotle, "Eudaimonia involves the recognition that one’s desire for the good is being fulfilled, and therefore one who attains eudaimonia is necessarily happy with his life." He explains how this sense answers to a desire whose satisfaction is itself part of eudaimonia:

We human beings...would never be able to attain our good with any regularity, unless we had effective desires for what we think worthwhile. Since we are creatures with strong desires for the good, as we variously conceive it, it is natural and inevitable for us to develop a deep interest in whether or not such desires are being satisfied. An animal with first-order desires, but no strong second-order interest in whether those first-order desires are being fulfilled, would not be fully human. Put otherwise: no person would choose a life in which he remains continually unaware of whether or not he possesses the good....So a major human good is the second-order good which consists in the perception that our major first-order desires are being satisfied. And

this second-order good is one we must have in order to be *eudaimon*.\textsuperscript{332}

There is something a bit misleading about this description, since any desire for an end is a desire for what *satisfies* it (as I have argued). To be distinct, the second-order desire Kraut mentions must be a desire to *know* that my first-order desires are and probably will continue to be satisfied. It is clear that the satisfaction of this second-order desire could be a major part of human happiness, as Kraut convincingly argues, but it is equally clear that this prospect cannot *itself* be what inspired the first-order desires. They must arise from my sense that their ends are either means to, or part of, my well-being, not from my sense that a *desire for them* is one I can knowingly satisfy. For if the latter were my reason, then what I desire is the satisfiable motivation to pursue certain first-order ends, rather these ends themselves: but in that case, I would have to *project* these ends in order to acquire the first-order motivation I hope to satisfy. Instead, we have here another form distinct from (5)-(6):

(7) The *eudaimon* cares about several worthwhile goals for their own sake.\textsuperscript{333}

(8) The *eudaimon*’s knowledge of his success in realizing these goals he pursues for their own sake contributes to his overall *eudaimonia*.

Like (5)-(6), form (7)-(8) is plausible, but likewise it cannot help us avoid the paradox for eudaimonistic theories of motivation. The agent’s recognition of (8) cannot explain the truth of (7) for him, because even though (8) fulfills his natural second-order desire, the goals mentioned in (7) are by hypothesis not pursued as a means to satisfying the second-order desire that (8) satisfies (and indeed it is only because of this that the successful pursuit of these goals *does* satisfy this second-order desire).

\textsuperscript{332}Ibid, pp.172-173.

\textsuperscript{333}Note that Kraut does not specify that ends we consider “good” or “worthwhile” must refer back to our own well-being: he leaves open the possibility of motivation pure of self-interest, as Annas also insisted.
—Friendship and Virtue

We reach the same result in considering Aristotle’s accounts of friendship and kindness. As Nancy Sherman argues in her valuable study, for Aristotle, "both friendship and goodwill require the non-instrumentality of our beneficence."

As she explains, self-sufficiency is a mark of eudaimonia for Aristotle, "But since friends are among the goods which make a life self-sufficient, self-sufficiency is relational and the good life is dependent upon and interwoven with others." To realize the benefit, again, we must value our friends for their own sakes. In the friendship relation, in particular, this means pursuing ends together, developing the "capacity to share and co-ordinate activities over an extended period of time." This kind of loyalty involves a kind of constraint on our acquisition of ends: "Ends are co-ordinated not merely within lives, but between lives. Thus, just as a particular choice I make is constrained by my wider system of objectives and ends, so too is it constrained by the ends of my friend." This means that our friendship involves the joint development of a conception of our joint eudaimonia. The "consensus" between friends does not simply precede and ground their relation, as it might for parties contracting for mutual self-interest:

In true friendship, we might say, friends realize shared ends which develop through the friendship and which come to be constitutive of it. Specific common interests are the product rather than a pre-condition of the relationship. Together my friend and I develop a love of Georgian houses, having had no real interest in them earlier.

---

335 Ibid, p.128.
338 As Sherman puts it, "One takes on, if you like, the project of a shared conception of eudaimonia" (p.133). And the eudaimonia so conceived is not my own only: the decisions I make are affected by my sense of their implications for my friend. So "what is relevant to the decision goes beyond the eudaimonia of a single, isolated individual" (p.134). My motivation is thus not egoistic at bottom.
339 Ibid, p.133.
I want to say that this is a perfect example of projective motivation, grounded in this case by considerations about what our friend thinks and feels. Let us imagine that Professor Sherman knows her friend has started to take an interest in the history, aesthetics, and characteristics of Georgian houses (perhaps after acquiring one). This ‘taking an interest’ on the friend’s part may itself be (partially or wholly) a projective act, or it may result from some desire. Sherman herself, I imagine, has no antecedent desire to learn about Georgian houses: the fact that her friend is taking an interest does not represent a sense on Sherman’s part of any lack of well-being or deficiency in her own life (if it did, this would be motivation by envy, competitiveness, or a comparative sense of inferiority rather than by friendship). Instead, in a kind of freedom from desire for it, Sherman projects the end of learning about Georgian houses, visiting them, etc. on the grounds that this will allow her to join in her friend’s new activities.

But, the response comes, doesn’t Sherman’s motivation instead derive from the more fundamental desire to pursue joint activities with her friend, which is part of her eudaimonia. At this point, several ways to work out this alternative to a projective explanation confront us, but none of them can succeed. First, we could say that pursuing joint activities is valued as a means to the terminal end of building the friendship, which the agent values intrinsically as part of her flourishing. This fails because it encounters the same paradox as before: if this were the agent’s real motive, she would not be being friendly. If I simply want to be the kind of person who has a friendship relation, because this is good for me, then will not be concerned for my friend’s needs and interests for their own sake in the very way necessary to have this relation. Second, as in the (3)-(4) model above, we could instead say that (i) the

---

340 Children encounter this problem sometimes when the set out too explicitly with the intention of making some friends, e.g. to alleviate loneliness. Approaching others in this needy fashion signals to their potential friends that this is a person who will not be committed to them purely or simply in the joy of projecting such a
agent desires ends she can pursue jointly with her friend for their own sake, which is what is necessary for the real sharing of pursuits constitutive of friendship; and (ii) she came to value these ends intrinsically because of recognizing (perhaps in deliberation with her friend) what realizing them would contribute to the good shape of her own life as a whole. This suggestion fails, because cases relevant for friendship do not conform to the (3)-(4) model: it is very unlikely, for example, that Sherman thinks becoming an expert on Georgian houses itself will improve her life very markedly, or that, even if she did think realizing this end would improve her flourishing, that this is why she came to value this end for its own sake. In friendship, as Sherman’s account makes so clear, it is not so much the particular goals, however independently valuable they may be, as the shared pursuit of them in joint activities, that is so intrinsically valuable.

Finally, an explanation following (a suitably modified version of) the (5)-(6) schema above would reflect this: (i) Sherman is noninstrumentally motivated to learn about Georgian houses, and (ii) pursuing this end in conjunction with her friend will contribute directly to her eudaimonia. This has a good chance of being right, but as I have argued, the truth of (ii) could not be the basis of a noninstrumental desire for the end mentioned in (i), because this end itself is not what is singled out as eudaimonistically valuable in (ii). So the motivation mentioned in (i) could only be projective.

So Annas and others are absolutely right that for Aristotle, virtue must involve certain ‘pure motives,’ but, contrary to the impression they give, his eudaimonistic moral psychology is inadequate to explain how these pure motives are possible. It is precisely because his analysis of the virtues correctly perceives the need for pure motives that it leads

340(...continued)

commitment, but instead is moved by a feeling of lack that he believes having a friend will redress. And this of course hinders the child in developing real friendships.
him beyond the resources of his desiderative theory of motivation and thus into a deep though rarely perceived inconsistency. Hence it is only *half-true* to say that "charges of egoism made against ancient ethical theories because of their eudaimonistic form miss the mark completely" and are "radically mistaken."\(^{341}\) Such theories are not egoistic in content, as Annas says, because they "give virtue a non-instrumental role in achieving happiness..."\(^{342}\) But this is the problem, since the dominant orexic conception of motivation in eudaimonistic theories, which is the *reason for* their formal structure centering on the agent’s own highest good, is inconsistent with the existence of just the sort of purely virtuous or moral motivation we find in the content of these theories. Thus modern critics are not entirely to blame for their confusions about ancient eudaimonistic ethics!

——*The Entry into Practices*

A similar conclusion might be drawn with regard to MacIntyre’s well-known treatment of goods internal to practices in *After Virtue*.\(^{343}\) MacIntyre begins with the notion of a "practice" as a "coherent and complex form of socially established cooperative human activity" that involves a certain kind of reflexivity: (1) such an activity is teleologically constituted in the sense that it involves "standards of excellence which are appropriate to, an partially definitive of, that form of activity;" and (2) the goods realized in trying to realize the ends of the practice in an excellent way, or "trying to achieve those standards of excellence," *include* the systematic "extension" or enrichment of its participants’ conception of the ends of the practice and the goods involved in devoting oneself to them.\(^{344}\) Thus,


\(^{342}\)Ibid, p.322.


\(^{344}\)Ibid, p.187. Note that this is my own way of breaking down and partially analyzing the definition of (continued...)
applying this analysis to one of MacIntyre’s examples, "Planting turnips is not a practice [but] farming is," because the former refers to an activity that is not reflexive in a way that the latter is: putting turnip-roots into the ground in a customary way is simply a behavior done with the intention of growing turnips, while "farming" is a more complex project that, if we devote ourselves to the end of farming well (and all this involves), will enrich us through a better understanding of what farming can achieve, what its possibilities are, and what joys it can bring quite apart from making a profit. Practices are reflexive in the sense that their aim is partly second-order: i.e. in conjunction with others (both in the past and the present) to redefine and extend the particular goals the practice can encompass: thus "practices never have a goal or goals fixed for all time."345

This reflexivity that distinguishes "practices" from simple or first-order activities depends on the phenomena which MacIntyre calls "goods internal to practices." His famous example of teaching the recalcitrant child to play chess offers a preliminary explanation of this notion:

Notice however that so long as it is the candy alone which provides the child with a good reason for playing chess, the child has no reason not to cheat and every reason to cheat, provided he or she can do it successfully. But, so we may hope, there will come a time when the child will find in those goods specific to chess, in the achievement of a certain highly particular kind of analytical skill, strategic imagination and competitive intensity, a new set of reasons, reasons now not just for winning on a particular occasion, but for trying to excel in whatever way the game of chess demands. Now if the child cheats, he or she will be defeating not me, but himself or herself.346

---

344(...continued)
"practice" which MacIntyre presents in a more compact form, but I believe it is true to his meaning.

345Ibid, p.193. MacIntyre refers to this reflexive feature to distinguish "practices" in his sense from technical skills "even when directed towards some unified purpose and even if the exercise of those skills can on occasion be valued for their own sake." Taken by itself, the employment of skills to some fixed end is only part of a practice; it lacks the historical sense of itself and reflexive ability to extend its purposes in light of this self-awareness necessary to constitute a practice.

346Ibid, p.188.
This is very plausible, but it is worth pressing the question: what exactly happens in this crucial transition from playing chess for the sake of getting candy-money to playing it ‘for its own sake.’ We might imagine that the transition occurs because the child recognizes that the internal goods MacIntyre describes are inherently enjoyable, or (more eudaimonistically) are essential components in a well-rounded life. But notice the problem with this suggestion: if it is just that, along the way, the child recognizes its incompleteness without these goods, conceives what I have called an S3-desire for the difficult strategic engagement that playing chess well provides, and plays chess for the sake of attaining that end, she will suffer from a different kind of self-defeat (similar to the way in which becoming reflectively aware of one’s focused concentration on something ruins the whole gestalt of that concentration).

The child has to play with winning fairly—or perhaps playing well—as the terminal end of her activity if she is to realize the internal goods; she cannot take these goods themselves as the goal she pursues for its own sake. And in general, goods internal to practices as MacIntyre conceives them will be goods that are realized only in pursuits that are not undertaken for the sake of realizing these goods.

MacIntyre recognizes that achieving the goods internal to (or constitutive of) practices requires a kind of pure motivation, in which eudaimonia is not the agent’s direct aim (since these goods are realized only in the collective pursuit of other ends for their own sake). He initially distinguishes external and internal goods as follows:

On the one hand there are those goods externally and contingently attached to chess-playing and to other practices by the accidents of social circumstance—in the case of the imaginary child candy, in the case of real adults such goods as prestige, status and money. There are always alternative ways for achieving such goods, and their achievement is never to be had only by engaging in some particular kind of practice. On the other hand, there are the goods internal to the practice of chess which cannot be had in any way but by playing chess or some other game of that
specific kind.\textsuperscript{347}

But "internality" to a practice does not rest primarily on that practice being the \textit{sole means} to a good. It has to do more directly with the way an internal good is realized. Someone aiming at an external good treats the practice \textit{merely} as a means, and thus does not really engage in it. Someone engaged in a practice realizes its internal goods by pursuing standards of excellence with their \textit{intrinsic worth} in mind: thus the authentic participant must pursue such excellence for its own sake. Only thereby does she achieve the related internal goods. Thus in MacIntyre’s interesting example of portrait painting as a practice, "There is first of all the excellence of the products, both the excellence in performance by the painters and that of each portrait itself."\textsuperscript{348} And it is by aiming to \textit{paint a good portrait}, and at the second-order level, aiming to interpret what should count as ‘good’ here, that the artist realizes a good in himself as agent:

...it is in participation in the attempts to sustain progress and to respond creatively to problems that the second kind of good internal to the practice of portrait painting is to be found. For what the artist discovers within the pursuit of excellence in portrait painting —and what it true of portrait painting is true of the practice of the fine arts in general— is the good of a certain kind of life.\textsuperscript{349}

Thus it is only via a sustained motivation that aims at goods (and the articulation of goods) \textit{separate} from the agent’s well-being as its final ends that the agent realizes what we might call the \textit{agent-goods internal to the practice} (MacIntyre’s second type). And, not suspecting the problem I described with reference to Annas above, MacIntyre links this feature of his analysis to Aristotle’s account: "As Aristotle says, the enjoyment of the activity and the enjoyment of the achievement are not the ends at which the agent aims, but the enjoyment

\textsuperscript{347}Ibid, p.188.
\textsuperscript{348}Ibid, p.189.
\textsuperscript{349}Ibid, pp.189-190.
supervenes upon the successful activity in such a way that the activity achieved and the activity enjoyed are one and the same state.\textsuperscript{350} This needs some modification to fit MacIntyre’s analysis of practices, since he shows that not only the \textit{successful} achievement but also the purified pursuit of excellences may also generate such enjoyment as an \textit{internal} agent-good involved in the practice.\textsuperscript{351}

But MacIntyre does not see that there is a much more serious difficulty here, because the fact that we can engage in practices with the kind of pure motivation needed to realize its internal goods is quite incompatible with the Aristotelian idea that all of a human agent’s motivation is governed (even if only \textit{formally}, as Annas urged) by the flourishing or eudaimonia of the agent as its final telos. If this broadly desiderative theory were correct, it would be impossible to see how someone could \textit{get into} a practice in the right way. The transition from the external to the internal attitude cannot \textit{itself} be motivated by the agent-goods internal to the practice without becoming self-stultifying. Only by \textit{projective motivation} can this transition occur. MacIntyre’s chess-playing child must at some point \textit{give himself} the goal of playing chess well as a terminal end, if he is to enter into the practice and eventually live its internal agent-goods. He may only recognize that he has done so later, and in retrospect, but the change involved requires projective commitment at some level, whether clearly introspected or not. Moreover, the child can have a variety of grounds for doing so, e.g. the inherently interesting nature of the challenge involved in chess, the chance to exercise his capacities, to form friendships, etc., but his grounds cannot be such as to generate S3-desires instead.

Theories of habituation are characteristically vague about this crucial transition precisely

\textsuperscript{350}Ibid, p.197.
\textsuperscript{351}As we will see in the next section, Frankfurt gives an analysis with the same implication.
in order to finesse the difficulty I have emphasized. If we imagine that this transition occurs by pure conditioning—that there arrives a moment when the chess-playing child is just habituated into playing chess for the challenge of playing well and possibly winning—we would have to accept the incredible thesis that no one can ever voluntarily enter into a practice in such a way as to realize its internal agent-goods. The emphasis on habituation in MacIntyre’s child chess player example still suggests the possibility of this analysis, which is one of its weaknesses. For although it is plausible in the case of training a young child to enjoy complex activities (like chess) that the child comes to acquire the pure motives that will be valuable to him quite involuntarily, due only to the initiative of others guiding him, this cannot be the true in every case of entering into practices in such a way as to realize their internal goods.

In this context, let us return to my earlier example of the value of pursuing a given philosophical question as grounds for projecting this end. It is true that participating in the practice of philosophical inquiry and cultivating its virtues may add to my eudaimonia. But my motivation to participate in the practice cannot be explained on this basis: instead, it must be rooted in my projection of the ends to which this practice is dedicated. This is true for all genuine practices in MacIntyre’s sense. Even if one’s initial motives for getting involved in a practice were ordinary ones like the desire to follow in a parent’s footsteps, or the desire to feel that one belongs to a group, one’s motivation for continuing involvement in the practice becomes projective in time, and this is the key to realizing the goods internal to that practice.

Let us take a case that will help distinguish this projective motivation relevant to practices from closely analogous S3-desires. Suppose someone becomes involved in a practice because she has S3-desires for the ends which the practice serve: e.g. someone
resolves on becoming a doctor out of a desire to reduce death and suffering in her community. This is an S3-desire because its depends on her visceral horror or aversion to needless loss of life and suffering, and her belief that she will flourish in the role of fighting these evils. And this motive is sufficient to ensure that she is involved in the practice of medicine and she values its internal goods intrinsically, as she would not if she entered medical school simply on the motive of gaining wealth (to satisfy S2-preferences). But the deeper motive for becoming or remaining a participant can either be an S3-desire or a projection. Suppose that in time, with success as a doctor, her own happiness starts to depend less on successfully reducing death and suffering but she nevertheless remains committed to practicing medicine. Then her motivation for participating in the practice has outlived the satisfaction of the original desire and become projective. She remains dedicated to the goal of medicine because she wills that it be her end.

Another person in the same community might abhor the same suffering, seeing it as wrong and needless, although he does not have any sense of visceral reaction to it or internal disturbance that can be returned to equilibrium only by realizing the goal, say, of making the community more healthy. He does not desire this goal, or experience it as something that draws him towards fulfilling a want in himself, but acts on thoughts such as: treating the sick would be a worthy enterprise; I have the opportunity to get into medical school now; and I have no other plans that conflict with this. On these grounds, he projects Hippocratic ends as his own goals, and thereby motivates himself to enter the practice of family medicine. In the process of pursuing this project, he eventually also develops S3-desires for the internal goods of his practice, but these desires are in Nagel’s sense ‘motivated.’ This shows how projections may motivate an agent’s desire for goods internal to a practice, without the agent’s initial motive for participating in the practice deriving solely from S3-desires for its
ends. If this were impossible, in fact, then the habituation necessary to acquire many of the most interesting and important S3-desires would (as Aristotle may have thought) be wholly dependent on fortunate intervention or benign manipulation by others, and could never be self-willed. Accepting this would be a high theoretical price to pay, especially since many moral conceptions of responsibility for character do presuppose that a responsible agent can regard herself as capable of self-reform and ordinary effort to develop her talents and various capacities, in better and worse ways, without depending entirely on others.

II.6 The Existential Significance of Projection

II.6.1. Frankfurt on Meaning and Final Ends

Several points still need to be made about projection to help distinguish it from desire. This section concerns the relation of projective motivation to existential meaning, or our sense of what gives life significance. The subsequent two sections concern the kinds of rational considerations and valuations which may ground projections. Finally, the points discussed in §7.3 will clarify the relation of projective motivation to desires that can accompany it. For our projections can take on a motivational life independent of prior desires with non-projective grounds that occasioned the projection; they can be grounded in nonconsequentialist ‘backward-looking’ considerations that also foster related desires; and they can work in tandem with desires to motivated the formation of particular intentions to act.

One distinguishing feature of projective motivation is the fact that our anticipated satisfaction in achieving the end is noncomprehensive for us, in the sense that by itself it does not explain the ‘value’ which the end and action towards this end (taken together) has for us. For the grounds which inform the ours self-projection onto an aim or object
—thereby making it our end— do not concern solely the value of achieving the end, but also concern attributes such as the worthiness of pursuing this end, or the importance of projection towards this end itself.

Harry Frankfurt draws attention to a phenomenon related closely to this aspect of projection in an intriguing essay "On the Usefulness of Final Ends." After briefly sketching the distinction of means and ends, and the associated distinction between instrumental and "terminal value" (i.e. "the desirability of the proposed object of choice apart from its usefulness as a means to other things"), Frankfurt argues that adherence to this scheme in Aristotelian moral psychology implies a "fundamental asymmetry:"

[First] A means derives its instrumental value from the relationship in which it stands to its end, but an end derives no value from the relationship between itself and the means to it.... [Second] A means derives no terminal value from being useful...Of course, what has instrumental value may have terminal value as well. But it cannot have the latter by virtue of the fact that it has the former.

Frankfurt believes that this Aristotelian approach is "impersonal," or "diverts attention from the fact that every end is the end of an agent," and as a result misconstrues how ends function and the complexity of their relation to means. Human agents need ends not just for their terminal value, but also so that their activities and thus their life can be meaningful for them. For a life to be "meaningful" in the first-personal sense, rather than simply in the sense of representing something for others, requires "final ends." Frankfurt says that while it is good for life to be meaningful, it is not the only important good, since a life that is

---

353 Ibid, p.3.
354 Ibid, p.4.
355 Ibid, p.5.
meaningful to its agent could still be unhappy, or evil.\textsuperscript{358} But this reflection on ‘existential meaningfulness’ —as I will call it— does show that final ends have an existential function \textit{beyond} picking out states of affairs that are intrinsically valuable or objectively desirable. The value of having final ends cannot be reduced to the value of achieving or realizing those ends:

...when \textit{is} activity important to a person? It is important to him only when he is devoted to something that he cares about. Thus a person’s life is meaningful only if he spends it, to some considerable extent, in activity that is devoted to things that he cares about. It is not essential that the activity he devotes to the things he cares about be successful. The extent to which life is meaningful depends less upon how much it accomplishes than upon how it is lived.\textsuperscript{359}

This point about the role final ends play in the meaningfulness of life helps clarify the possibility of the theoretically unfamiliar (though common) kind of motivation I am calling ‘projection.’ When an agent projects an end rather than desiring it, this agent’s motivation arises not so much from an attraction to a perceived value the end has for his life, but from caring about the very activity of pursuing this end as part of what makes his life meaningful. Of course, a person’s cares may by misguided or corrupt, but it is crucial that when an evil motivation is ‘projective,’ it cannot be explained merely as the result of a concupiscent desire for something the agent perceives as his good (as in Thomistic models). Rather, the agent projects a deviant end because he cares about engaging in some activity \textit{despite} recognizing its immorality, or its detrimental effects on his long-term well-being; he can positively define the meaningfulness of his life in terms evil pursuits without having to be motivated by S3-desires in doing so.

Of course this argument depends on the intuition that \textit{meaningfulness} is not a fixed end like eudaimonia; it cannot be desired like human flourishing, not only because an increase in

\textsuperscript{358}Ibid, p.7.
\textsuperscript{359}Ibid, p.8.
meaningfulness can make us more unhappy, but also because to have a meaningful life in the relevant sense is to have given it meaning. The ‘need’ for meaning is thus not really a desire that gets satisfied: it is an open-ended eagerness to use our capacity for projection, a will to make something our business, to involve ourselves actively in life, or to develop commitments that will both be informed by and help shape what a sense of what is important to us. On the basis of this role it has in making life meaningful, Frankfurt argues that the existential value of a final end exceeds its ‘terminal value’ as an end:

Final ends are possible states of affairs, which someone values for their own sakes. It must not be supposed that the measure of how a life is lived is given by the value of his final ends. Rather, how a life is lived is a function of what it is like for the person to pursue them. The problem of selecting final ends is not the same, then, as the problem of measuring the inherent or terminal value of possible states of affairs.\footnote{Ibid, p.8.}

We might add that as a result, an agent’s selection of final ends cannot be determined simply by her S3-desires for them: rather, her selection must involve some projection of these ends guided by attention to their existential value or the meaningfulness of pursuing them.\footnote{Like "existential meaningfulness," the terminology of "existential value" is also mine, rather than Frankfurt’s. As the title of his article indicates, Frankfurt seems to think of this analysis as showing that final ends have a kind of "instrumental value" for making life meaningful: "our final ends derive a certain instrumental value from the very fact that they are terminally valuable," and thus make certain means instrumentally valuable (p.14). I think this gives the wrong impression, since it might lead one to think that final ends have instrumental value as a means to satisfy the end of having a meaningful life, which is valued intrinsically. And this suggests that meaning is just another end we desire. Instead, what Frankfurt really means to show, I believe, is that the very structure of means-ends explanation is inadequate to capture every kind of human motivation—which is precisely what the possibility of ‘projection’ also implies.}

As Frankfurt puts it, "The goals that it would be most desirable to achieve are not necessarily those it would be best to seek." This is not simply for instrumental reasons, such as that some desirable ends are impractical. It is because there are essential differences "in the kinds of activities, and in the patterns of activities, by which various final ends may be
pursued.”362 And different activities by which ends are pursued may vary in how intrinsically interesting, worthwhile, challenging, or (overall) how meaningful they are.

Thus activities, Frankfurt argues, themselves possess a kind of intrinsic value which affects the existential value of the ends for which they are the means:

...since living a meaningful life is important to us for its own sake, useful activity possesses for us not merely instrumental value but terminal value as well. Being engaged in the pursuit of a desirable state of affairs is not desirable exclusively because it is desirable that the state of affairs should obtain.363

This shows that the relation between means and ends is not as asymmetric or unidirectional as Aristotle thought. As Frankfurt explains, the point is not just that “activities may be desired as final ends and not merely as means to ends other than themselves” —a point which Aristotle himself emphasizes. Of course activities themselves can be desired as ends because of their "intrinsic character." But "Aristotle does not recognize that [activities] may possess terminal value precisely because they are instrumentally valuable"364 —or because they are instrumental for pursuing ends whose realization is valuable for its own sake.

Frankfurt’s point could be expressed by saying that there is a unique form of terminal value that attaches to the very pursuit of ends itself. This is the intrinsic existential value that useful activities have independently of the value of achieving their ends. Existential value in this sense is unlike the terminal value which activities can have for Aristotle when they are constitutive of a terminal good. For even those activities that are only means to such goods, rather than an intrinsic part of the human good, still have intrinsic value as means, because pursuing ends (especially in interesting and challenging ways) has terminal existential value above and beyond the terminal eudaimonistic value of the ends pursued.

This reading suggests that the underlying reason why the value of having final ends must be distinguished from the value for well-being of realizing these ends themselves, as Frankfurt has indicated, is that there are different sorts of terminal value, which are correlated with fundamentally different kinds of motivation. When ends are projected rather than desired, the activities involved in pursuing these ends acquire an existential kind of terminal value, which is unlike the intrinsic value for the sake of which objects or states are desired as ends. It is the possibility of projective motivation, which is especially relevant to the deepest projects of a person’s life through which its unity and ultimate meaning are decided, that explains why a person’s motives are not completely determined by the perceived ‘desirability’ of realizing ends.

This insight extends beyond Frankfurt’s basic point that people want useful work not only for its products, but because "useful work is among their final ends." In addition, it implies that determination of which pursuits are worthwhile or important involves kinds of valuation other than those which motivate by desire: for example, Frankfurt suggests that "Pursuing one final end rather than another may lead a person to engage in activities that in themselves are more enjoyable," which depends on their variety, complexity and capacity to give unity of meaning or purposefulness to someone’s life (perhaps by requiring full development of their talents, conforming to their emotional temperament, and so on). But ultimately, on my diagnosis, it is because our activities are projectively motivated pursuits of ends which we give ourselves that they acquire lasting existential value as activities. For our projections constitute the existential or first-personal meaningfulness of our lives: they are grounded in part by our sense that some ends and activities in pursuit of them are important.

365Ibid, p.15.
366Ibid, p.15.
to care about, independently of their power to satisfy preferences or contribute to our eudaimonia, and in part by our sense of the importance of caring itself. As Frankfurt says in an earlier paper,

It is manifest that the varieties of being concerned or dedicated, and of loving, are important to us quite apart from any antecedent capacities for affecting us which what we care about may have. This is not particularly because caring makes us susceptible to certain additional gratifications and disappointments. It is primarily because it serves to connect us actively to our lives in ways which are creative of ourselves and which expose us to distinctive possibilities for necessity and freedom.367

The distinctively non-teleological structure of projective motivation allows us to understand these features of caring and existential meaningfulness to which Frankfurt has drawn attention.

II.6.2. The ‘Good’ of Projection vs the Grounds of Projection

As I have already suggested, the possibility of projection implies that there are ways of valuing ends and their pursuits which do not simply foster desire for these ends. Frankfurt’s analysis thus does not imply that people engage in projection just as a means to satisfy a ‘desire’ to have meaning in their life. As Viktor Frankl has argued, the "will to meaning" is a "primary concern" of persons.368 Though Frankl speaks of "man’s desire for a life that is as meaningful as possible,"369 calls such meaning a "higher need,"370 and argues (with compelling examples from his own experience in the Holocaust) that having meaning and purposes can be critical for mental health and even survival, this should not be taken literally

---

367Frankfurt, "The Importance of What We Care About," in The Importance of What We Care About, p.93.
370Frankl, The Unheard Cry for Meaning, p.33.
as affirming that projects are undertaken by persons on the motive of gaining fulfillment or happiness.\textsuperscript{371} For as Frankl himself insists, the person often becomes his "true self" only through concerns that are directed to persons and causes outside himself:

He becomes so, not by focusing on his self-actualization, but by forgetting himself and giving himself, by overlooking himself and focusing outward....What is called self-actualization is, and must remain, the unintended effect of self-transcendence; it is ruinous to make it the target of intention...It is the very pursuit of happiness that obviates happiness.\textsuperscript{372}

"Self-actualization" is a state that depends pursuing other objects as ends-in-themselves; it is not itself usefully conceived as an intrinsic goal. Similarly, meaningfulness achieved in projects and positive endeavors contributes to our human well-being principally through our ends being projected for their own sake, and not for the sake of eudaimonia (and this apparent paradox seem to be part of Frankfurt’s point as well).\textsuperscript{373} Thus we can recognize that projecting worthwhile ends may contribute to flourishing, and hence that there is a ‘good’ of projecting (namely the ‘good’ of one’s life having meaning for one); but unlike Aristotle’s human goods, this good can only be realized through acting on motives other than the desire for it, i.e. by taking up or projecting ends for reasons distinct from their ‘good’ for the agent. It is essential to projection that it not be grounded (solely) in any eudaimonistic value-judgment, including one concerning the good of projection. If the value of projection for existential meaning is to be realized, it can at most be a ‘foreseen but

---

\textsuperscript{371} Rather, it is more a reflection of the fact that our language, after 2500 years of influence by traditions stemming from Greek moral psychology, pushes us towards portraying all motivation as desiderative.

\textsuperscript{372} Ibid, p.35-36, my italics. In a recent presentation ("Eschatological Ultimacy and the Best Possible Hereafter") I describe a similar (and well-known) paradox in the relation between ethics and eschatological meaning: the hope for a final or ultimate validation of ethical endeavor cannot be its motivating intention, or it fails to be ethical endeavor. Since eschatological hope is closely related to the problem of existential meaningfulness, the similarity between these paradoxes is not accidental.

\textsuperscript{373} Compare this to Frankfurt’s remark in his essay, "The Importance of What We Care About," that there is something liberating in the experience of being "seized" or "captivated" by an object of our love, through commitment to which we move out of ourselves (The Importance of What We Care About, p.89).
unintended’ consequence; or more accurately, it can be intended, but this value cannot be the comprehensive explanation for the agent’s motivation, as I put it earlier.

The relation between projective motivation and existential meaningfulness is thus not usefully explained or analyzed in terms of ‘desire.’ If projections were motivated by the simple desire to have ‘some meaning or other,’ then they would be able to satisfy such a desire by projecting almost any end, and would therefore be arbitrary. But projective motivation cannot be arbitrary: if the agent can cease pursuing her end or reverse her decision on a whim, she lacks commitment or resolve, and hence turns out not to project this end after all.

Furthermore, projections, like S3 desires, must always have "reasons" or grounds, as I will call them. Grounds in this sense do not motivate projection, since ‘projection’ names a nonteleological structure or mode of motivation itself; rather, we could say that such reasons or grounds projectively motivate us to pursue some end. If there is such a thing as projective motivation, then there must be reasons for caring about ends and their pursuit other than any ‘goodness’ they possess in the eudaimonistic or teleological sense. For projection is a type of motivation which responds to other kinds of evaluative considerations than judgments that something is ‘good’ in the sense being able to close a gap or to requite an incompleteness in one’s happiness. The reasons which grounds projective motivation instead spur us to take up some goal as our cause, or to make something our business, though it need not have been.

Note that by "reasons" for action, I mean grounds for motivation that could be the cognitive content of individual practical deliberation, and sometimes also interpersonal debate (though some of the thoughts underlying projections may be too private and inarticulable to meet this discursive condition). This does not mean that these grounds must
be ‘rational’ in any ideal sense: they can be full of error, self-deceptive, or even perverse, as can the projections which arise from them.

Sometimes philosophers speak of any motivational state, such as desires and perhaps also emotions, as "reasons" for an intention or action, because they explain why the intention was formed or acted on. In this sense, "reasons" is just a synonym for "motives." By contrast, "reasons" as rational grounds lie behind and figure in the agent’s coming to have the motives she does: they explain rather than constitute her motivation. If motivation is susceptible to internal explanation (as opposed to external explanations that would undermine the motivational state if the agent were aware of their role, such as physiological and unconscious causes), i.e. if motivation can be rationally influenced in conscious thought, then there must be agent-accessible reasons or grounds for her motives which do not themselves constitute motivational states in her.

There is some potential for confusion here, since writers in contemporary action theory sometimes use "desire" and "motive" just to mean any considerations in the background of act in terms of which we can understand it as an action. For example, Justin Gosling says that "when we talk of someone doing something for a reason, this is just another way of saying that they did it because they wanted whatever it is that is specified in the description of the reason."374 In these terms, the ground or reason why we devoted ourselves to something will be thought of as our "motive" for doing so.375 But a motive in this logical sense is not necessarily a motive in the psychological sense, i.e. a conative state constituting an inclination to change the world in some way. Although by definition it would be a

---

375 And if we also use "desire" in the thin sense, where is means any kind of motive, then reason and desire will be intertranslatable as well: in this framework, it becomes a truism (rather than a psychological explanation) to say "that rational action is action done from some desire" (Gosling, p.174).
motive in the logical sense, the ground of a projected state of motivation need not be a motive in this psychological sense until we let it be. (And even if the ground itself is a psychologically motivating state, the motivation created by projecting ends for ourselves on its basis is distinct from that furnished by its initial grounds, and may even outlive its occasioning conditions, as we will see).

II.7. Kant and Projective Motivation

II.7.1. The Limitations of Deontic Projection

It is easy to see that what is called ‘motivation by specifically moral reasons’ in the Kantian tradition fits the S4 structure of projection. As Kant says in Theorem III of the second Critique, "The material [or end] of a practical principle is the object of the will. The object either is the determining ground of the will or it is not." Projection is precisely a kind of motivation that is not determined by our prior apprehension and evaluation of the end projected. This does not mean that for Kant a "practical law," or maxim that has a universalizable form, has no end or object at all: "Now it is certainly undeniable that every volition must have an object and therefore a material, but the material cannot be supposed, for this reason, to be the determining ground and condition of the maxim." Without this point, there is no way to understand Kant’s later discussion of the "Highest Good" as the end or object of the will when it is determined by pure practical reason:

The moral law is the sole motive of the pure will...Consequently, though the highest good may be the entire object of pure practical reason, i.e. of a pure will, it is still not to be taken as the motive of the pure will; the moral law alone must be seen as the ground for making the highest good and its realization or promotion the


377 Ibid, §8, Theorem 4, Remark I, Ak.35, p.35 (Beck’s italics omitted).
object of the pure will.\textsuperscript{378}

In short, the highest good is a ‘projected’ rather than desired end in moral motivation. And in general, while it is true that for Kant, without "reference to outcome, to results, to materia, to ends," there can be "no willing,"\textsuperscript{379} this does not mean that motivation is ‘teleological’ for Kant in any substantive sense. Intentional action must always be goal-directed, but the \textit{adoption} of the action-maxim in our second-order agency involves projection of its goal, and it may be projected on grounds \textit{other} than the goal’s value for personal happiness.\textsuperscript{380}

Underlying this view is a conception of the will which, though not very clearly articulated, radically resists older teleological conceptions of motivation.\textsuperscript{381} Agents form intentions through their "will, i.e. [a] faculty of determining their causality through the representation of a rule" or "according to principles."\textsuperscript{382} These intentions will always have a form such as ‘Do A in order that X,’ i.e. they will have an end, yet desire for the end may not always be the \textit{ground} on which the agent adopts or wills the intention as the expression of some rule or principle. Often, as in maxims following from the principle of happiness, "an object of choice is made the basis of the rule and therefore must precede it."\textsuperscript{383} Yet even such heteronomous acts require an intention that has a minimally rational cognitive form,

\textsuperscript{378}Ibid, "The Dialectic of Pure Practical Reason," Ak.109, p.115.


\textsuperscript{380}Thus Watson is not quite right, I think, to infer from Kant’s remark that happiness is necessarily the desire of every finite rational being that "Willing per se is necessarily involved with happiness" (\textit{Tradition(s)}, p.106). Kant identifies volition with our second-order agency, which need not decide to intend ends because of their value for the agent’s well-being (even if they have such value). Nor does the inevitable \textit{reference} to ends in the second-order intentions of decisions themselves imply that they are moved by concern for happiness, although it \textit{does} imply that even decisions to adopt maxims on moral grounds must have an end or goal, with reference to which we can define \textit{a noneudaimonistic} notion of the ‘Good’ that \textit{follows from} the ideal of the ‘Right.’

\textsuperscript{381}This resistance is apparent in the opening claim of the \textit{Groundwork} which, as Otfried Höffe argues, implies that "In contrast to traditional moral philosophy, absolute goodness does not consist in a supreme object of the will...such as Aristotle’s happiness, but in the good will itself" (\textit{Immanuel Kant}, p.142).


and are therefore not simply caused by inclinations or impulses that lie behind them:

the will is never determined directly by the object and our representation of it; rather the will is a faculty for making a rule of reason the motive of an action that can make an object real.\(^{384}\)

Since Kant distinguishes practical "rules" in general from "laws" (rules with lawlike form) and will from "pure will" (which is motivated only by laws), it is clear that this remark applies not only to action from the motive of duty, but to heteronomous acts as well.\(^{385}\) The will is thus coextensive with the domain of spontaneity, which, as Allison says, "concerns rational agency in general, that is, the capacity to determine oneself to act on the basis of general principles (whether moral or prudential)."\(^{386}\) As several Kant scholars have recently emphasized, this notion of volition reflects Kant’s unorthodox view that desires and inclinations do not by themselves even constitute *motives* to act, but only *grounds* for possible motives:\(^{387}\) only when the will gives them the discursive form of *reasons* or principles do they acquire motivating force.\(^{388}\) In his recent review of Barbara Herman’s work, Paul Guyer summarizes this point as follows:

On an empiricist [i.e. Hobbesian/Humean] conception of agency, feelings, desires, and other naturally occurring and empirically given conditions are, as such,


\(^{385}\)Of course, some Kant scholars uses "laws" rather than rules for relating ends and means in maxims. Thus Otfried Höffe says that for Kant "The capacity to act according to representations of laws [rules] is also called the will, so practical reason is simply the capacity to will"— see Höffe, *Immanuel Kant*, tr. Marshall Farrier (SUNY Press, 1994), p.139.

\(^{386}\)Henry E. Allison, "Spontaneity and Autonomy in Kant’s Conception of the Self," in *The Modern Subject: Conceptions of the Self in Classical German Philosophy*, ed. Karl Ameriks and Dieter Sturma (SUNY Press, 1995): 11-30, p.11. The original version of this paper was presented at the *Conference on German Idealism*, University of Notre Dame, April 8, 1994), and the essay is reprinted in Allison’s collection, *Idealism and Freedom* (Cambridge University Press, 1996).

\(^{387}\)As Allison says, according to Kant’s Incorporation Principle, which covers "actions motivated by inclination as well as purely moral considerations," “although a finite rational agent is still sensuously or ‘pathologically’ affected, that is to say, it finds itself with a set of given inclinations and desires, which provide possible motives or reasons to act, it is not causally necessitated to act on the basis of any of them” (Ibid, p.13).

motives, functioning as causes for action. So here the idea of action in the absence of feelings or desires does not make any sense to begin with, and actions can instead be expected to arise only from an agent’s strongest feelings or desires in any particular situation....Kant’s conception of moral worth seems paradoxical, Herman suggests, because we approach it with such a picture in mind, but Kant’s conception of agency is not at all like this. For Kant, no naturally occurring and empirically given condition such as a desire or inclination is ever itself a motive for action; a motive reflects an agent’s "reasons for acting," or his decision, for example, to accept a desire as an adequate ground for action in light of his underlying principles and other relevant circumstances...  

This gives part of Kant’s reasons for holding that heteronomous action is also spontaneous or free in the transcendental sense, since it is never directly determined by desire. As Charles Nussbaum says, "Though heteronomous and conditioned, empirical practical reason is spontaneous, even if not, strictly speaking, autonomous." And as Allison argues, according to Kant it is essential to the very concept of a self that 

To think of oneself as a rational agent requires presupposing that one is capable of projecting ends, acting on the basis of self-imposed general principles (maxims) and in light of objectively valid norms.

For Kant, this implies that what counts as a motive for us is underdetermined by incentives: "It is not that one’s desires are irrelevant to the determination of what one chooses to do (they are obviously the source of reasons to act); it is rather than they are not sufficient reasons." Or as we might better put it, they only become practical reasons or motives for us through spontaneous incorporation. As Barbara Herman explains, 

Kantian motives are neither desires nor causes. An agent’s motives reflect his

---


390 Charles Nussbaum, "Kant’s Changing Conception of the Causality of the Will," International Philosophical Quarterly, Vol. 36.3 (September 1996): 265-285, p.269. Nussbaum emphasizes that the same problems of understanding how the will can be noumenally free and yet affect a closed and determined causal order arise for the empirical will as for the pure will on Kant’s account.

391 Henry Allison, "Kant on Freedom: A Reply to My Critics," in Idealism and Freedom, p.126. Allison’s use of the term "projection" here is convenient and suggestive, but I should note that it was not the original basis for my own choice of this description for non-teleological motivation.

392 Ibid.
reasons for acting. An agent may take the presence of a desire to give him a reason, as he may also find reasons in his passions, principles, or practical interests. All of these, in themselves, are ‘incentives’ (Triebfedern), not motives, to action. It is the mark of a rational agent that incentives determine the will only as they are taken up into an agent’s maxim.\(^{393}\)

In other words, Kant’s position is really that all genuine motivation is projective in my sense. This is a reasonable explication of his view, although of course he does not explicitly formulate the notion of a projective motivation as I have here. But so interpreted, his position is more extreme than I endorse, since it arguably implies that animals with quite sophisticated desires that lack "will" in Kant’s sense cannot be motivated to act.\(^{394}\) Kant’s analysis may also imply that all genuine motivation is second-order, or consists in acts of incorporation that refer to other potential motivating states, such as inclinations. My analysis denies this, and holds that projection can be a first-order state alongside others.\(^{395}\)

But Kant’s view that only projection is motivating it is based on an important insight: namely, that various types of incentive (which I treat as other types of genuine motivation), such as desires, inclinations, emotions, and other states, may themselves serve as the grounds for projection. For example, when someone desires to obtain a college degree in order to

---


\(^{394}\) This follows unless the conditions for motivation are different in lower animals, which might be a possible Kantian response. Then Kant’s thesis would be that in any rational being, i.e. their negative practical freedom or power of spontaneously determining motives by incorporating grounds into maxims ensures prevents unincorporated incentives that would count as motivating animals from motivating such a rational being. This position, which begins a line of thought leading all the way to Sartre, implies that we can never be unfreely motivated, or moved against our will, and runs up against difficult counterexamples in psychology. Such a view likely results from linking motivation very strongly to action, so that nothing is a motive unless it entails an act or attempt to act on it. Then the idea that a rational being cannot act on a motive without an act of ‘consenting’ to it (consent which we are able to give or refuse), gets transferred back to the very formation of the motive itself: a ground must itself have consent to become a motive. It is more plausible to hold that we can acquire motives without consent, though action on them may always require some type of consent by the agent.

\(^{395}\) But as we will see in Chapter IV, I do recognize a special role in the formation of a coherent self for projections with second-order structures, i.e. projections that commit the agent to other psychological states (such as desires, impulses, or first-order projections) as her end. Such projections are centrally involved in a person’s own autonomous control over her motivational character, which is what concerns Kant most.
have more opportunities to get higher-paying jobs, the fact that this end drawn her in a desire-mode (S2) may by itself, or (more likely) in combination with other considerations, give her grounds to project this end as well. She doesn’t merely desire it, but also thinks something like ‘yeah, I will get that degree, by golly, because I’m perfectly right to want a good job’ (perhaps adding ‘since I deserve it,’ or ‘since I’ve been waiting for this for a long time,’ etc.) and through resolve projects graduation as an end for herself. Then on my account, she is overdetermined with two different types of motive to pursue this end (though on Kant’s account, she only has one real motive, namely the principle by which she makes the desire for a job her reason for studying hard to pass her courses). Yet since this insight does not necessarily entail Kant’s more extreme position that only projections count as motives at all, the insight can be retained in my account without taking that additional step.

If this is right, then the idea of projective motivation permeates Kant’s account of moral agency. As Otfried Höffe summarizes, "Practical reason, as Kant says for short, means the capacity to choose one’s action independent of sensible determinations such as instinct, desires, passions, and sensations of pleasure and pain." But interpretative care is needed here, for there are two different senses in which actions might be motivated ‘independently’ of impulses arising from our sensuous nature as an empirical being.

In the first (weak) sense, as we have seen, even heteronomous acts are ‘independent,’ since they make impulses into motives only by incorporating them voluntarily in forming intentions or maxims on their basis, or projecting their satisfaction as the agent’s end. In

396 Höffe, Immanuel Kant, p.139.

397 A projective structure entails that the motivation is ‘weakly independent’ in this sense, which Kant (in my view unpersuasively) extends to all human motivation. It also entails that the motivation is ‘strongly independent’ of desire or inclination in Kant’s sense, and to this extent projective motivation is autonomous. But it does not entail autonomy in Kant’s special sense of impartial motivation, because the considerations which ground projections may contingent to the person, or not motivated by their universalizability, even if they are in fact universalizable.
the second sense, which corresponds to Kant’s notion of auton-omy, to be ‘independent’ requires that the action-maxim or first-order intention not be derived even indirectly from impulses or desires for happiness arising from our sensuous nature. Yet even in this second stronger sense, as Henry Allison has pointed out, it seems that actions whose motive is neither derived from desires nor from strictly moral considerations could be ‘strongly independent’ of sensuous impulses. Therefore Kant’s implicit assumption that autonomy or strong independence occurs only in the form of deontic motivation, or adoption of a maxim because of its universalizable form, is unjustified. Yet this assumption is arguably the main basis for Kant’s derivation of the categorical imperative in the Groundwork Part I, and for his argument by dilemma in the Second Critique that since the end or object of an action-maxim cannot determine the will in moral motivation, only the lawlike form of the maxim can determine the motive in this case:

Since the material of the practical law, i.e. the object of the maxim, cannot be given except empirically, and since a free will must be independent of empirical conditions (i.e. those belonging to the world of sense), and yet be determinable, a free will must find its ground of determination in the law, but independently of the material of the law. But besides the latter there is nothing in a law except the legislative form. Therefore, the legislative form, in so far as it is contained in the maxim, is the only thing which can constitute a determining ground of the [free] will.

When this implicit (though unjustified) exclusion of non-formal but strongly independent motives is combined with Kant’s tendency to underemphasize the distinction between weak

---

398 As Allison has argued, autonomy means "independence" not only in the sense represented by "the capacity of the will to determine itself to act on the basis of self-imposed principles (which would include heteronomous principles)" but also in the sense that the will "recognize[s] sufficient reasons to act that do not stem (even indirectly) from its needs as a sensuous being," as do the reasons made sufficient in heteronomous maxims or intentions (see Allison, "Spontaneity and Autonomy," in The Modern Subject, p.18).


400 On this topic, see my paper, "From Autonomy to Morality: The Most Important Gap in Kant’s Deduction of the Categorical Imperative" (currently unpublished).

401 Ibid, §6, Problem II, Ak.30, p.29.
and strong independence itself, the result is to give the impression that only moral
motivation in his formal sense is even weakly independent of desires, or has a projective
structure. Thus many readers have acquired the impression that for Kant, autonomy means
motivation by pure practical reason, and only autonomous motivation (in this formal
sense) is ‘projective’ in the sense of not be motivated by prior desires for the ends of one’s
actions. Although this is not the correct reading, it accounts for the frequent
misunderstanding that only action done for the sake of duty counts as "free" in any sense for
Kant.402

Let us call formal moral motivation in Kant’s sense ‘deontic projection.’ Although
"autonomy" is a distinct concept for him, on Kant’s account, such formal or deontic
projective motivation exhausts autonomy. On my account, however, although deontic
projection constitutes one important kind of projective motivation, it certainly does not
exhaust the range of S4-states. Thus, without explicitly describing it, Kant detects the
possibility of projection, but he links its structure with the derivation of a formal principle of
morality, and thus he suggests that the paradigm case of projection is the self-giving of an
end on the ground that the form of one’s maxim to act for this end fits a universalizability
criterion (i.e. the ‘contradiction in willing’ test). This focus prevented Kant from exploring
in any detail the manifold other ways that considerations which are neither mere desires nor
moral considerations in his strict formal sense can supply grounds for projecting ends.
Therefore the philosophical world did not learn from Kant about projection as a broad type

402 As Allison notes, "heteronomy" is frequently misunderstood as "as simple lack of agency, a complete
subjection of the will to the ‘causality of nature’" ("Spontaneity and Autonomy," in The Modern Subject, p.17).
In §1.10.6, I will briefly discuss Kant’s own link between his notion of the possibility of moral motivation and
our alleged noumenal libertarian freedom in the form of transcendental spontaneity.

403 For example, deontic projection may figure prominently in citizens’ having a "sense of justice" in Rawls’s
sense. For citizens must project the end of building basic political institutions that would count as substantively
equal and fair from an impartial perspective, if they are to be committed to "justice" in the deontological sense.
of motivation with diverse subcategories that are *neither teleological nor essentially deontic*, because it did not learn that deontic motivation is hardly the only way to acquire a motive without being moved merely by desire for the ‘object’ or ‘material’ of the maxim. In eliding the possibility of autonomous yet non-formal and non-desiderative motivation, Kant obscured the full potentiality of his break with the teleological paradigms of antiquity.

II.7.2. Non-Deontic Grounds for Projection

What, then, are these grounds for projection which Kant missed? My earlier examples suggested that the class of reasons that could provide grounds for projection include (a) *aesthetic reasons*, such as that the pursuit of some end is varied in its challenges, or fits one’s talents, or is enjoyable; and (b) *ethical reasons* of many nonformal varieties, such as the ideal of neighbor-love,\(^{404}\) or the intuition that an end is important to humanity, or the habituated sense that its pursuit is virtuous. (c) We should also include a class of *already-motivational* reasons, such as the fact that I have inclinations or preferences (S2), or that I desire something (S3) because I evaluate it as good for me. As Kant feared, the latter category may account for a large part of the projections which are irrational (if stemming from distorted evaluations), immoral, or unwise. (d) In addition, a sense that we have an *real opportunity* to pursue some cause or relationship-goal may figure among our reasons for projecting it as our end (though such a pragmatic sense of opportunity is perhaps rarely by itself a sufficient reason for projection).\(^{405}\)

Sometimes, however, even in the beginning projections do not stem from simple desires or emotions, nor apparently from anything we could plausibly interpret as aesthetic, ethical,
or moral judgments, or the pressure of a temporary ‘window of opportunity’ for some meaningful endeavor. Thus we must allow for another class, (e) which I will call *personal reasons*, for lack of a better term. These include both quasi-fatalistic reasons that can hardly be articulated, such as that I was called or *meant* to pursue this project, that someone or something is *right* for me, and the inchoate sense that pursuing a certain end expresses what I want to be. In a few instances, personal reasons may seem to radically underdetermine projection, and yet be their only apparent ground: we may verge on simply committing ourselves to project a certain end just because we can, or because we decide to give rein to our power of projection itself.

In other instances, personal reasons might even be radically evil, including (e.g.) the idea that pursuing a given end is wrong, shocking, or rebelliously bold — considerations that ground some of the most perverse projections of which human beings are capable. In particular, projections made on entirely malicious grounds are possible even when the agent knows that pursuing such an end is diametrically contrary to his own well-being, happiness, or overall satisfaction, and such cases offer some of the strongest phenomenological evidence for the independence of projection from the eudaimonistic/teleological structure of motivation which has dominated the history of western philosophy to the point of covering over the existence of alternatives (hence the interest existentialist authors often show in examining radical evil).

Unfortunately, however, a detailed phenomenology investigating what different types of grounds for projection there may be and how they might operate has not been worked out, largely because Kant’s approach has led philosophers to think of his deontic projection as the *only* conceivable alternative to traditional teleological motivations, whether of the Platonic/Aristotelian S3 variety or the Humean S2 type. Thus the diverse realm of
possibilities comprehended under the basic motivational structure of projection has remained largely uncharted, since philosophers, under Kant’s influence, have assumed that any motivation supposedly transcending the teleological structure of desire could only be what Habermas calls "the weak force" of impartial reasons, with its (frequently alleged) unappealing abstractness.

II.7.3. Separating Projection from Accompanying Desires

Among other grounds, I have suggested that experiencing motivational states such as an emotion or desire for some end X can themselves facilitate an agent’s projection of X. It might be objected that this is impossible unless the resulting projective motivation is just really the same desire for X: otherwise, are we not multiplying entities unnecessarily? In fact, however, there are cases where we must acknowledge that desiderative motives precede and occasion projective motivation, leading to a contingent overdetermination of motives. This is implied by the fact that projections can also be separated from their occasioning desires in such cases.

Imagine someone who desires to enrich himself, and believes the best means to this end is to start a criminal ring. He dreams of living a life of peaceful luxury and he takes this as a reason to pursue wealth by establishing an effective and ruthless crime ring. But as time goes on, other motivations also are at work. Eventually, he projects the success of his crime ring, and his pursuit of criminal activities, as a way of giving meaning to his life—even though he did not originally desire these activities for their own sake. The work of building his gang did not itself make him happy, and its maintenance remains laborious for him, but he comes to be dedicated to it in a way he never initially imagined. Because this projection of criminal ends is motivationally distinct from the desires for wealth that overshadowed it
at the outset, these motives can *coexist* in uneasy symbiosis, or one can remain while the other withers.

For instance, imagine that our subject has attained a vast wealth, and his desire for riches is sated; or his original dream of living in a Caribbean paradise fades and comes to appear adolescent and uninteresting to him in his old age. After attaining the status of ‘godfather,’ our criminal kingpin loses his initial desire for simple luxury. Yet, notably, he remains more committed than ever to developing the criminal ring’s influence. He is no longer moved by a desire for contentment and sensual pleasures, since now what he ‘really wants’ is to be involved in the day-to-day manipulation of people, exercise of power, the thrill of the kill, and so on. The thought of retirement no longer enters in, but he is bent on pursuing his criminal practice. We might say that what was formerly a mere *means* to his desired end has taken on a life of its own for him. This is a common experience, which occurs in many different variations, yet it is often not very usefully explained just by hypothesizing a new terminal desire for what was formerly just the agent’s means to an end. Rather, the agent has come to value the evils internal his corrupt practice by *projecting* these activities and their immediate ends as his or her goals: projective motivation has detached itself from the desires out of which it developed.

This kind of case illustrates how goals that are adopted initially because they are means to satisfy pre-existing desires have the tendency to become *projected* ends-in-themselves, pursued because of the meaningfulness of the pursuit itself. Because projection is a motivation distinct from desires that may occasion or prepare the way for it, it can take on a life of its own. This helps explain the difficulty people often have explaining why they are so committed to the practices in which they are engaged, or why they may simply respond, ‘because this is what I do.’ It may also help explain the related (but perhaps somewhat
different) sense that one has ‘lost the point of it all’ somewhere along the way. This may be a case of misdiagnosis: when people discover that they are pursuing ends they no longer desire, they often assume they are caught in a mindless routine, when in fact the explanation may be that they have projected these ends for themselves.⁴⁰⁶ Or it may be a correct diagnosis when a person finds that her persistence in some set of activities is controlled by desires quite distinct from her original projective commitment to an end which these activities are supposed to serve: e.g. a police detective who joined the force out of love for his community finds that he now lives only for the thrill of the chase, the glory of big catches, or the vindictive pleasures of interrogation. This policeman has lost his original projected commitment, and is now moved only by S2-desires. Yet for this very reason, in an analysis of moral worth, we would probably have to rate his condition as the lesser evil relative to the godfather’s passion, which shifted from merely shallow and selfish desires to an active and resolute self-investment in corruption for its own sake.

Of course, such motivational shifts can also occur in more fortunate ways. I may adopt the end of becoming someone’s friend on the dubious grounds that I want to sleep with her and appearing friendly will increase my chances; but in the process, the initial desire may disappear (or be altered) while my motivation to spend time with her, help her out, talk over ideas and plans, etc. remains, and without any sense that it is necessary for my well-being. Rather, I now project the end of her happiness for its own sake. And interestingly, I think the latter projection can grow out of the former desire without requiring a clear formulation of some new justifying ground other than the fact that I took on the role of being friendly to her in the past. This suggests that familiarity may itself be a consideration that can motivate projectively; as in the judicial argument of stare decisis, for projection, that something was

⁴⁰⁶ Thus they should really say that they ‘found the point of it all’ along the way.
sought in the past may itself be a reason (if not often the sole or controlling reason) for projecting it now.

This takes us to another feature of projection which distinguishes it from desires, at least of the S1-S2 types, namely, its potential to be grounded by what Elizabeth Anderson calls "backward-looking" reasons: "agents often choose an alternative not because it maximizes expected future payoffs but because the alternative bears an appropriate relation of narrative unity to prior action." Anderson uses this feature to show that her own principle of respect for persons recommends ends for different reasons than alleged consequentialist equivalents that would recommend the same ends. For instance, a forward-looking consequentialist principle that "fulfilling [past] commitments to oneself is intrinsically good" recommends maximizing the state of affairs that commitments are kept, and so would encourage people to "make commitments willy-nilly, just so that more commitment-fulfillments can exist in the world." Moreover, on any consequentialist principle recommending the bringing about some state of affairs E, the motive for pursuing E is always such that "there can be no rationale for avoiding E-violations when they would bring about more instances of E." Thus no principle R defining what is rational or good in terms of "agent-centered restrictions" can be captured by a consequentialist putative equivalent calling on us to maximize the state of affairs that R is realized.

If this is right, then there must be nonconsequentialist motives for action that do not motivate us to promote ends, but rather to pursue ends in certain appropriate ways. Nor can

---


408Ibid, p.542.


the motive in such cases simply be to maximize instances of end-pursuits guided by these restrictions. For example, as Anderson insightfully points out, if

I take my own desire to be a loving person as my reason for action...[then] I do not act out of love for my husband....What makes sense of my taking my husband’s needs as my reasons for action is that I love him, not that in so taking them I make myself a loving person.411

Thus the motive of love would not motivate her to harm her husband if this would make her more loving in the long-run. And the structure of this motivation is projective: it is not a desire for the objective good or value that being loving arguably has for the agent.

Similarly, backward-looking motivation to fulfill past commitments cannot be understood simply as a desire for the state of affairs that one’s commitments are fulfilled. I infer from this that only when an end is projected can the grounds of our motivation in pursuing the end be genuinely backward-looking in Anderson’s sense. Thus the agent’s own history and past commitments provide another type of personal reason for projections.

Of course, the idea that pursuit of an end fits with one’s own narrative development and provides personal continuity might also be cited as a reason to desire achievement of that end in the S3-sense, i.e. because our well-being or eudaimonia depends on maintaining narrative unity of character. But although it is not strictly consequentialist (since it does not seek to maximize a state of affairs), the motivation in such a desire is still essentially forward-looking, as any S3 state must be. What I ‘desire’ in this case is not the end itself, but the part of my eudaimonia that is constituted by keeping faith with my past commitments by pursuing this end (whether actually realizing it would add to my well-being or not). Yet this would surely be an unusual desire, which seems very similar to the case in which I actually project the end, whether or not I desire its realization, simply because it fits or is

411Ibid, p.545.
implied by my past commitments. Could this be because the idea that pursuit of an end (as distinct from realizing the good constituted by some activity) contributes to human flourishing simply stretches the notion of objective well-being too far? What maintaining the narrative unity of my life through fulfilling commitments and making my purposes coherent over time seems to realize is not so much my happiness, but rather first-personal meaningfulness of my life, as Frankfurt’s deconstruction of the means-end relation suggested. Pace the usual eudaimonistic versions of virtue ethics, it is possible to build such narrative unity and meaningfulness around commitment to morally evil or socially warped projects that cannot promote human flourishing, or (more commonly) to build it around fine projects without the resulting narrative unity dramatically adding to one’s happiness or fulfillment in life. Indeed, the requirements of narrative unity that I feel compelled to maintain may lead me to project ends that are so heavy a burden on me that I know my attempt to pursue them all will doom me to many dissatisfactions —yet I project them on these personal grounds anyway. These reflections suggest that backward-looking reasons for action will normally be grounds for projective motives rather than forward-looking S3-desires, though such desires with backward-looking reasons are (just) conceivable.

It is important to add that projection can supply the motive for the same intentions to act that are often formed on the basis of S3-desires as motives. For example, returning to an earlier case, I may desire an education because I recognize that it will be good for me both instrumentally (in increasing my earning power) and intrinsically, in improving my self-esteem and opening me to new possibilities. But I may instead (or in addition) project the end of becoming an educated person because I think the attempt is admirable whether it succeeds, or because it provides useful work to do in time that I would otherwise waste, or because I project another end which becoming educated serves, or several of these taken
together. Projection may partially overlap desires, or it may overlap with one of the agent’s desires against others, thus aligning the agent with the former against the latter. Introducing the notion of projection thus helps us come to grips better with the real complexity of human motivation, which is its normal state rather than the exception.

At this point, we can draw the conclusion towards which the last two sections have been leading: agent-initiated ‘projective motivation’ is what people unschooled in philosophy and its technical definitions ordinarily mean by will. As I said at the outset, our everyday prephilosophical notion of will involves not only the notion of ‘deciding to act’ (or second-order agency) but also the idea of ‘resolve,’ or a decision with a special motivational quality. In the phenomena of projection, we have disclosed the basis for this thick notion of volition as a phenomenon of the ‘middle-soul.’ This preliminary analysis of projection gives us an outline of the distinctive meaning that volition has alongside other middle-soul phenomena, such as S3-desires, emotions, and evaluative attitudes. While the classical account of the will as something between lower appetites and speculative reason —as ‘intellectual appetite’ or desire informed in some way by evaluative reasoning— took us in the right direction, it was not sufficiently discriminating to pick out the will by itself. This classical definition tended to obscure the relevant differences between S3 and S4 states and to collapse them into one category. Our preliminary phenomenologies of desire and projection have shown how to avoid this conflation. A more complete analysis would also analyze the complex relations between volitions as projective states and the many varied types of emotion, in order to clarify the distinction between will and passions.  

II.8. Projection and Libertarian Freedom: A Concluding Note

412 I have pursued this issue in other work in progress, but because of the complexity of contemporary literature on emotions, space considerations precluded including this analysis in the dissertation.
I have argued that the *ability to will* that is distinctive of persons is the capacity for second-order agency through which persons not only form intentions to perform first-order acts, but do so on the basis of a projective kind of motivation that is not desiderative or incentival, nor simply a function of practical reasoning and its evaluations. Given the existentialist aim of this thesis, it might look as if my analysis is meant to suggest that projection means *freely* giving oneself a practical end in some suitably libertarian sense of ‘freedom.’ In this concluding section, I will briefly address this issue and forestall some potential misinterpretations.

There are two possible misunderstandings here. First, one might think that the distinction between desire and projective motivation works by contrasting the formation of desires, which is often held to be involuntary, with the formation of non-desiderative commitments, which is supposed to be a spontaneous creation of a motivation in oneself, as if ‘out of nothing.’ This misconstrual would have the effect of making projective motivation look both arbitrary and mysterious. Against this, I have emphasized that we always project ends for ourselves upon grounds that do not themselves depend for their status as *grounds for us* on the very same projections. Contrary to the familiar Humean picture, we recognize the importance of various sorts of considerations —deontic, informally ethical, aesthetic, and personal or existential— that transcend our existing set of motives, and the projection of ends is never without some rationale or basis of these sorts. The distinction between desires and projected motives is recognizable *in* the kind of motivation produced, and does not depend solely on *how* its emergence can be explained in terms of prior reasons or grounds: the phenomenal quality of the motivation itself varies because of the different agent-content direction of fit in the two cases. Thus this distinction would in principle be compatible with an intentional account of a person’s motives according to which, given the grounds she had
Throughout, I use the term ‘liberty’ as a shorthand for libertarian freedom or the capacity to bring about alternate possible options in the same choice circumstance.

Second, one might wrongly assume that the distinction I have drawn between types of motives depends ultimately on their having different origins in the causal sense, or on postulating some other sort of deep metaphysical difference between them. For example, in line with a recent trend, one might hold that a projected motive is distinct from a desire (and from other types of motivating emotional states) because it is agent-caused, whereas desires and passions typically are not. Or a libertarian who does not explain how free choices count as self-determined by recourse to a concept of agent-causation may still hold that in the case of projective motivation, the agent could have motivated herself otherwise, but not in the cases of desire or emotion (however he explains the choice to project an end as ‘coming from’ the agent). But neither of these hypotheses is required by my phenomenological analysis, which (as it stands so far) remains compatible with the possibility that, even though we might never know how, our projections could be as determined as all our other motives are by the initial state of the universe along with the laws of physics and neurochemistry.

I believe that further phenomenological analysis of reflexive projections concerning our own motivational make-up (see Chapter IV) and the conditions on responsibility for character would help ground the additional hypothesis that will in its distinctive sense as projective motivation generally extends to alternative volitionally possible projections in any given choice-circumstance, and that this liberty of the will distinguishes it both from rational deliberation per se (which may not motivate) and from desiderative motivation. Others have already taken steps in this direction. For example, Pink argues that the main reason for accepting second-order agency is to recognize the ordinary intuition that we have

\[\text{Throughout, I use the term ‘liberty’ as a shorthand for libertarian freedom or the capacity to bring about alternate possible options in the same choice circumstance.}\]
alternate-possibilities freedom not only of action, but also of the will or decision-making. This the freedom of the will is transferred to the level of action: we "always retain a continuing freedom not to act as we have decided," because "decision-making is a method of future action control which is essentially non-manipulative," or freedom-preserving. Thus Pink holds that the will is not a deliberative but an "executive agency," since the intentions that it forms

...are quite distinct from practical judgments. One can form intentions which are not accompanied by any conviction that what is intended is desirable. One can form intentions which are not actually explained by concomitant practical judgments at all. Nevertheless, the capacities for practical judgment and intention are still possessed together, because the core function of intention formation is to apply or execute our practical judgments about how we should act.

This fact that one "can decide acratically" goes along with the fact that "It is very much more intuitive that we have freedom of decision than that we have a freedom of practical judgment." Similarly, in his own less sophisticated but interesting treatment, the neo-Hegelian Rudolph Steiner holds that the most important question about freedom cannot even be framed unless we talk in terms of second-order agency and its motives for forming intentions:

Should it matter to me whether I can do a thing or not, if I am forced by the motive to do it? The immediate question is not whether I can or cannot do a thing when a motive has influenced me [freedom of action] but whether only such motives exist as affect me with compelling necessity...The question is not whether I can carry out

415 Pink, p.6.
416 Pink, p.29.
417 Pink, p.23. Keep in mind that Pink uses "desire" in a generic, formally open sense. He does not offer a theory of motivation distinguishing between "desire" in various substantive senses and other types of motivation. So for him, acting against your most rational 'desire' can only be akratic. On my account, by contrast, projections come in both akratic and non-akratic varieties.
418 Pink, pp.30-31.
a decision once made, but *how the decision arises within me.*

On this basis, Steiner is skeptical (for example) of Eduard von Hartmann’s model, which holds that the motives for our decisions follow from the interaction of our environment with the individual "characterological disposition" that makes us who we are. Hartman’s theory is a precursor of contemporary psychological models of action as a function of ‘personality type’ (clusters of dispositional traits) along with situational variables. All such models leave us no liberty in the determination of our motives, whether or not we enjoy liberty in forming specific intentions and acting on them.

But if the will in the thick sense —second-order agency that forms intentions through projecting their ends— is free in the libertarian sense, this still requires a separate argument going beyond the initial phenomenological analysis and classification given in this chapter. And such an argument will have to contend with the objection that if the will not only makes decisions about intentions but also forms the *motive* for such second-order acts, adding libertarian freedom at this level would result in a complete ideal of "autonomy" that will be inexplicable without recourse to arbitrary regress-stopping choices. While I have tried to forestall this objection by introducing the idea of *grounds* for projective motivation (rooted in such aspects of personhood as the agent’s historical acquisition of character and volitional dispositions, her moral reasoning, and her orientation to existential meaningfulness), it remains to be shown that grounds sufficient to make projections non-arbitrary are compatible with libertarian freedom of the will. In the meantime, as a quote Steiner provides from Robert Hamerling indicates, it is possible to hold that volition both involves second-order agency and the projection of motives for decisions, without believing in its

---

On this point, let me return briefly to Kant, who held that human motivation is projective and that it is free in a libertarian sense that would be impossible if temporal nature as a closed causal order were the whole of reality. In a recent critique of novel reinterpretations by Henry Allison and Allen Wood, Karl Ameriks has argued that, although "Allison does a good job of showing how Kant himself was attached to this conception" of agent-freedom, neither Kant nor Allison provide sufficient grounds for this conception against compatibilist alternatives.

I will briefly evaluate three aspects of Ameriks’s argument in light of the foregoing analysis of projective motivation. First, I'll consider his argument against Allison’s view (following Kant) that the conception of ourselves as agents who act for reasons requires us to conceive our acts as involving libertarian freedom. Second, I'll consider Ameriks’s reasons for holding, pace Allision, that even if it could be established, the practical necessity to postulate libertarian freedom cannot stand by itself without prior metaphysical conditions (which for Kant are provided by a metaphysical rather than merely epistemic version of transcendental idealism). Third, I’ll consider Ameriks’s argument that actions which seem to us not to have desiderative motives could in fact still be determined by desires. While I’m largely sympathetic to Ameriks’s position on all three issues, I’ll raise several questions about each.

(1) Ameriks maintains that Kant’s own arguments for liberty of the will is required by the

---

420Steiner, p.41: "Should freedom of the will consist in being able to will something without reason, without a motive? But what does it mean to will something, other than to have a reason to do or strive for this rather than that? To will something without a reason, without a motive, would mean to will something without willing it. The concept of will is inseparable from that of motive. Without a motive to determine it, the will is an empty ability" (quoting Hamerling’s Atomistik des Willens [Hamburg, 1891, 2 vols., p.213]).

Thus Ameriks agrees with Allison about

Kant’s basic belief that, even if humans do universally seek pleasure, they are still rational and in some sense spontaneous agents who do so because of maxims that they have a ground for in a non-mechanistic sense. We can even agree, as Allison also stresses, that Kant believes these maxims have an expression in our empirical, temporal being, and that, as Kant says, this expression takes the form of ‘laws of freedom’ (A802/B830) e.g. about what ought to be done even for the sake of prudence. But we need not agree that the mere capacity to act according to such laws already defines us as having free agency in an incompatibilist sense.

In my terms, Ameriks’s point is that being moved by giving oneself maxims as opposed to being moved directly (or ‘brutum’) by desires and impulses is to be ‘spontaneous’ only in the sense of being motivated projectively, which does not necessarily imply that one’s projections are at liberty in an "incompatibilist" sense. To say that "on the Kantian theory of agency, one’s maxims must always be not sheer givens but ‘taken as one’s own’,” is only to say that maxims are adopted through second-order agency, whose acts themselves have the minimal rational form of intending to decide on such-and-such maxim for particular reasons. Thus a general capacity to adopt maxims (or project motives) does not demonstrate our libertarian freedom. The notion of projection helps clarify the vital distinction between these two different senses of ‘spontaneity.’

Since this critique, Allison has acknowledged that "the practical necessity of acting under the idea of [libertarian] freedom leaves in place the epistemic possibility that I am deluded in

\footnote{Ibid, p.220. Ameriks’s skepticism here depends in part on his view that the approach Kant takes in the \textit{Critique of Practical Reason} after his ‘great reversal’ is something of a "regression to dogmatism" since the epistemic possibility remains that the connection between moral duty and liberty is illusory.}

\footnote{Ibid, p.224.}

\footnote{Ibid, p.40.}

\footnote{This is why, in my paper on "The Meaning of Kierkegaard’s Choice Between the Aesthetic and the Ethical," I described the Kantian principle which Allison calls the "Incorporation Thesis" as equivalent to Frankfurt’s idea that for our will to be free, we must \textit{identify} with our action-maxim, or decide for our own reasons to adopt the intention of our first-order act (see pp.80-81).}
believing that I am acting," and this is a doubt that "cannot be exorcised by any theoretical means."426 But though this correctly reflects Kant’s view that the practical perspective does itself provide us with metaphysical knowledge, it does not answer Ameriks’s point, which is rather that it is not clear why the idea of oneself as a rational agent capable of adopting maxims also requires one to think of oneself as free in an incompatibilist (rather than merely projective) sense of ‘spontaneity.’ As Allison acknowledged in an earlier essay,

Why, after all, should the capacity to project ends, act on the basis of reasons, or even independently of desire, require the assumption of a mysterious contracausal faculty? Could it not rather be the outcome of a cognitive process that is completely explicable in naturalistic terms?427

His response is that libertarian freedom is not "introduced in an attempt to provide something like the best explanation for the ‘phenomenon’ of rational agency," but rather is conceptually essential to the idea of an agent who deliberates and chooses maxims.428 Whether or not it is true, this idea requires not only that our decisions are projectively independent of "passions or overwhelming urges" but also that they are "not...merely causal consequences of our antecedent states."429 But since Allison does not derive this additional condition from our incorporative capacity per se, he leaves it unclear why it should be vital to our sense of ourselves as persons.

As a result, we must concede Ameriks’s point that Allison has not shown that the conditions of moral agency require us to regard ourselves as free in the libertarian sense; like Kant, Allison has not given "adequate attention to the possibility of a sophisticated

---

compatibilism.” But there might be a promising way of reviving the neo-Kantian position on this issue by appealing to the epistemology of modality: in order for the significance of moral necessity to be accessible to us, we must be able to apprehend the significance of counterfactual possibilities of action and motivation. Since the meaning of moral necessity depends on seeing some volitionally possible motives as required, some as permitted, and some as not permitted in the circumstances of action, the meaning of moral necessity (which is given in experience) would be unintelligible to us unless we were able to see both moral and non-moral motivations as volitionally possible for us in a given circumstance, which is to see ourselves as enjoying a form of libertarian freedom. To see these options are only as physically or nomologically possible would be insufficient to provide a basis for our understanding of moral possibility and necessity, which concerns volitional possibilities, possibilities of our motivation and action.

Along these lines, one might still be able to defend a deduction of liberty from the ‘faktum’ of our access to moral modality, or the significance of moral necessity for us. Such a defense is not part of my project in this dissertation, but note that it would still leave open the possibility that our experience of moral necessity itself is an illusion.

(2) For this reason, concerns arise about Allison’s neo-Kantian position that we can focus on how persons who take themselves to be moral agents must regard themselves, without having to know that their actions and motives really are—or even can be—undetermined. Ameriks argues convincingly that for Kant, the practical requirement to regard ourselves as having liberty would not be enough if theoretical metaphysics did not at least imply the real possibility of undetermined agent-causation: "any actual determining must itself be regarded ultimately as either caused naturally or not so caused. If nature is the closed system that

---

430 Ameriks, "Kant and Hegel on Freedom," p.226. In particular, Ameriks thinks Kant failed to respond adequately to the Leibnizian, Wolffian, and Spinozistic compatibilisms of his day.
Allison allows, then the latter option appears excluded, even if, as he claims, we ‘can’t help’ but think ourselves free ‘regulatively’ in a practical respect. Thus for Kant, a practical requirement to regard ourselves as having liberty can only find a grip if we can independently establish an ontological "two-aspect" view with a "timeless" noumenal ground for our freedom. A remark in a more recent paper clarifies this point:

Kant properly indicates that, at the common sense level, even positive unanimity [about our liberty] from the practical perspective could not save the claim of our freedom if it were to come into conflict with what can be established about nature, either by science or metaphysics (KrV Bxxviii-xxix); in this one respect there is precisely not a "primacy of practical reason."....The doctrine of transcendental idealism is especially significant for him here because it means not merely that there might be some "non-natural" room for uncaused causing; it shows that there always must be some such "room," given the antinomies that arise from taking determined spatio-temporal nature to exhaust reality.

This seems right as an interpretation of Kant, and it also seems right more generally that, at least as long as we believe in knowledge of nature that is not ultimately dependent on our practical ability to regard ourselves as agents, any alleged practical requirement to regard...

---

431 Ibid, p.222.

432 Ibid, p.222. Allison also uses the term "two-aspect view" for his interpretation, but he means by this only that there are two epistemic perspectives, "two 'points of view' or 'ways of considering'" ("Reply to My Critics," in Idealism and Freedom, p.128), rather than two real aspects in the metaphysical sense (which is still distinct, as Ameriks insists, from a full two-worlds or two-entities metaphysics).


434 And this may be where several post-Kantian German idealists departed radically from Kant. They are followed in recent continental philosophy by a line of thinkers from Buber and Levinas to Nancy and Derrida who have accepted something close to Allison’s reading of the primacy of practical reason and our liberty as justified by their necessity for the very "conception of ourselves as moral agents" (Ibid, p.221). Unlike both Allison and Ameriks, these continental writers avoid the problem of a noumenal basis for such liberty by taking the practical perspective opened by the ‘faktum’ of moral responsibility to the Other as the sole ultimate perspective for our reason, the basis of a new purely practical ‘metaphysics’ in which interhuman relations and the moral unsubsumability of individual difference replace all theoretical judgments and ego-based apprehensions as the root of ‘knowledge’ (thus radicalizing Kant’s notion of the primacy of the practical). For these writers, then, the ‘faktum’ of the Other and our responsibility for her functions like an ‘intellectual intuition,’ but is not a theoretical intuition of the ego at all (instead, this egological model of knowledge as cognitive theoretical appropriation in which our noetic acts construct our noemata is precisely what is transcended). Ameriks follows Kant in rejecting such an absolute primacy of the practical, but this issue remains open and disputable.
our actions and motives as involving liberty depends for its force on its consistency with the 
results of our metaphysics of nature.\footnote{And this in turn requires either that we reject the closed deterministic picture of physical nature, or that we accept a dual-aspect view in which a non-physical and potentially non-determined aspect underlies the physical aspect of persons.}  

(3) But in responding to Allison, Ameriks makes a further argument which would seem to 
challenge Kant’s evidence that regarding ourselves as moral agents implies accepting our 
capacity for projective motivation (whether or not this would also require regarding 
ourselves as free in an incompatibilist sense).\footnote{Or at any rate, Ameriks’s argument here at least seems to obscure the difference between these issues.} He says that Kant invalidly infers 
from the \textit{psychological} absence of a particular causal \textit{content} in one’s intentions 
(i.e. one doesn’t see that one is acting as the ‘mere’ effect of a particular force [or desire]) to the \textit{metaphysical} absence of any natural cause as the \textit{efficient} ground of the act which has that content. Similarly, in moral contexts, the fact that certain 
maxims involve a rule whose content makes no essential reference to human desires still does not show that the actual adoption of such maxims...is not in fact 
caused by desires... The compatibilist hypothesis is not that ‘rules’ about sensible ends, or even beliefs about such rules, need be what cause our actions, but rather 
just that there can be desires present which in some, perhaps totally hidden, way are 
their ultimate efficient cause.\footnote{Ameriks, "Kant and Hegel on Freedom: Two New Interpretations," p.224, quoting Allison, \textit{Kant’s Theory of Freedom.}, p.225. In addition to the \textit{non-sequiter} described here, Ameriks thinks that Kant’s mature position also suffers from assuming too uncritically that the conditions of moral responsibility require alternate possibilities of action. See Ameriks, "Reinhold’s Challenge: Systematic Philosophy for the Public," March 18, 1998 draft, pp.23-24.}  

Two things are suggested here. First, even if heteronomous and autonomous motivation 
appear to be alternatives at the level of projective second-order agency or \textit{will}, the choice of 
either alternative as experienced may still be determined by hidden causes. "The 
‘phenomenal’ fact that the maxims permeating our moral life appear as ‘self-imposed’ rules 
hardly settles the question of whether this imposition is free of natural determinism."\footnote{Ameriks, "Kant and Hegel on Freedom," p.226.} Second, our projective motives might have \textit{desires} as their hidden causes. 

I agree with the first of these suggestions, but not the second. My analysis has not shown
that projection must involve liberty; it also does not deal with the question of whether
projections can have unconscious causes. It is possible that the agent’s full grounds for her
projections may only be imperfectly introspectable, although deciding to intend a first-order
act is itself something done intentionally, or for reason(s), and therefore the agent must
always (however unreflectively) have something ‘in mind’ in projecting ends—something
which makes her projection voluntary under one set of descriptions (or intensional
designations) and not others.439

But I think we can rule out the idea that apparently projective motives could have desires
or impulses in Kant’s sense as their underlying ground, cause, or motive (Ameriks uses these
terms loosely and interchangeably). Desires, even if they are unconscious, must have some
psychological content (as my analysis in §I.6-I.7 showed), and they may also be realized in
some brain state or other noumenal structure. It is possible that this underlying structure
could cause a projection, since we have not ruled out hidden causes, but this would leave the
"desire" realized in it only an epiphenomenal role. To avoid this, we would have to
 hypothesize that the content of the unconscious desire, rather than its underlying structure,
caused or grounded a projective motive. But this could only mean one of two things. (A) It
could mean that the desire for an end E included the projective motivation to pursue E, but
this is contradictory. (B) Or it could mean that the desire for E occasioned a new motivation
beyond the desire—one that could potentially survive if the desire were extinguished. This
seems possible in particular instances, but if all projections were occasioned by desires in
this way, it would be mysterious why we often experience motivation by desires that do not

439Thus while Ameriks reads Kant’s distinction between ends and grounds of an action to mean that
"whatever the ends are which make up the overt content of our intentions, our actions can also have hidden
grounds of a different form" (p.228), I instead read this distinction in terms of the two orders of agency and
intention. The end as Kant uses the term concerns the intended goal of a first-order action, whereas the ground
refers to the goal intended in the second-order intention on which the agent acts in deciding to adopt the first-
order intention.
prompt any extra projection or separable motivation to pursue the desired end. This suggests that when a desire does ground a projection, it must not by itself necessitate the projection: unless there were a ‘leap’ from desire to projection in these cases, would not every desire entail a corresponding projection in every case of human motivation?

Moreover, an existentialist need not conceive the psychological grounds of projective decisions as their "efficient cause:" one can hold that second-order acts are simply uncaused (and such a rejection of the agent-cause model does not imply that decisions are not actions). As Allison suggests, the idea that our decisions and acts are not "causal consequences of our antecedent states" does not "deny that there is any connection between an agent’s antecedent condition or underlying character and the ensuing action, but it is to deny that the connection can be understood in strictly causal terms." Choice need not be indifferent for it to involve a ‘leap.’ Of course, Ameriks can admit that motives may not be "caused" in this nomological sense, but hold open the possibility that they can be determined in some non-physical way by noumenal grounds, e.g. divine decree.

We have not ruled this out nor proven that it is incompatible with moral responsibility. Ultimately, I believe that the conditions of moral responsibility for one’s volitional character do require the hypothesis that we enjoy libertarian freedom at least in a certain kind of projective motivation —namely reflexive projection of ends concerning our own motives.

---

440 See the novel argument for such a libertarian alternative to agent-causation in Stewart Goetz’s "Libertarian Choice," Faith and Philosophy, Vol. 14.2 (April, 1997): 195-211. Goetz defends the thesis that "choice is essentially an uncaused mental action done for a reason, purpose, or telos" (p.195). He relies in part on Frankfurt’s point in a famous critique of causal models of action that "It is an epistemological feature of an agent who knows that he is making a choice that he knows this while he is choosing;" yet on causal accounts, he "cannot possess this knowledge in virtue of his awareness of the choice itself," but only by virtue of his awareness of its causes could he tell that it is a choice rather than another kind of event (these being distinguished only causally on such accounts) (p.198). Goetz also incisively defends the idea that choices can be actions with reason that underdetermine rather than cause them, without this making choices ‘random’ in any relevant sense (p.1991-200).

But this requires a separate argument that lies beyond the scope of this dissertation. It is worth noting, however, that even if the compatibilist were right that moral responsibility for a given act did not require anything more than that its maxim and end is projected, which is insufficient by itself to ensure that it was free in the libertarian sense, there might still be other reasons for positing libertarian freedom. For example, the fact that we can be responsible in this way for both good and evil actions itself surely entails liberty in one important sense in Kant’s scheme. The thesis that we can be responsible for adopting maxims on both on autonomous and heteronomous motives suggests a kind of liberty both in deciding which action-maxims (first-order intentions) to adopt, and what reasons to adopt them for (e.g. even an act required by duty may be adopted for self-interested reasons): we always at least have the real alternative of heteronomous or autonomous motivation. Kant’s position in the Second Critique reflects this: he "tries to show that our absolute freedom must be clear from (and solely from) our commitment to maxims that involve (either through acceptance or rejection) specifically moral ends." Ameriks is right that the projection of either kind of motive by itself does not imply liberty, but the phenomenological evidence that we can generally project either kind in the same choice-circumstance does lend support to the thesis that human moral agents must regard themselves as at liberty to act on different motives (although their basis for so regarding themselves may not lie solely in intuitions concerning the conditions of moral responsibility).

*     *     *     *     *

In closing, this chapter has provided a basic outline for an existentialist conception of the will. It has defended the thick concept of will by introducing projective structures as an interpretation of the distinct motivational role of volitions, and outlined the differences

---

442 Ameriks, p.225, my italics for emphasis. Ameriks’s point in this passage is focused on Kant’s derivation of freedom from moral capacities, but his wording suggests acknowledgment of the point I am making here.
between projections and other motivational states of human psychology. Although this analysis has lent the idea of projective motivation some initial plausibility, it remains to be shown that ‘projection’ in my sense can find any support (outside of Kant) the philosophical tradition, and to substantiate it by showing that there are recognized phenomena in human experience which can only be explained in terms of projection. The next two chapters will complete these tasks. Chapter III will show that the idea of projective motivation has existed in form (if not in name) at least since medieval times, and remains a recurrent theme in both continental and analytic discussions of moral psychology. Chapter IV will then show that the recognized (though controversial) phenomena of "volitional identification," as Frankfurt has introduced them, cannot adequately be understood except as forms of projective motivation with a special importance for the determination of moral character. Finally, Chapter V will examine and critique several alternative explanations of volitional identification.
CHAPTER III

Illustrations of Volition as Projective Motivation:

Frankfurt, Scotus, Levinas, and Arendt
III.1. Introduction: Five Key Points on Projection

III.2. Frankfurtian Care as Projective Motivation
   III.2.1. Caring as Challenging Oneself
   III.2.2. Care and Self-Unification over Time
   III.2.3. Care and Love

III.3. Williams’s Neo-Humean Internalism

III.4. Duns Scotus on Free Will

III.5. Levinas on Superabundant Will and Volitional Generosity

III.6. Arendt’s Distinction Between Circular and Linear Motivation
III.1. Introduction: Five Key Points on Projection

In Chapter II, through reflection on themes in Kant, Frankfurt, and some other examples, I briefly introduced five key points about projective motivation:

(1) In projection, the satisfaction of achieving the intended end is *noncomprehensive* in the sense that the considerations or reasons that motivate projection of the end are not focused solely (if at all) on the end’s power to satisfy the agent’s preferences, or on the objective value of the end for the agent’s well-being. The agent who realizes an end she projects of course experiences some satisfaction, but only *because* she has projected this goal for herself: this potential for satisfaction is derivative, and hence noncomprehensive.

(2) Projection depends on the fact that there are ways of valuing both ends and their pursuit quite aside from the value of the ends or activities—or even of projecting these ends—for the satisfaction of one’s prior preferences or the fulfillment of one’s personal well-being. Moral motivation in the deontic sense is one such mode, but there are many others. Personal reasons for projecting ends can include broader convictions that something is worth caring about (nonformal ethical convictions), religious reasons (e.g. ‘callings’), aesthetic reasons, or considerations related to the question of having worthwhile labor or meaning in life, none of which are strictly speaking part of *eudaimonia*.

(3) Projective motivation may often develop alongside or grow out of pre-existing desires, but it is separable from them and therefore can exist stably on its own after the precipitating or facilitating desires have faded.

(4) Projection can be motivated by backward-looking reasons, whereas all desires, including S3 desires based on objective eudaimonistic evaluations, are essentially forward-looking in orientation. Projection motivated by considerations of narrative unity between past and future commitments and by goods internal to practices looks to existential meaningfulness rather than to personal well-being as its grounds, *even if* such commitments do in fact contribute to *eudaimonia*.

(5) The same intentions could be motivated either by desires or projective grounds, or in some cases, by both. Thus there can be cases of overdetermined motivation, in which projection aligns the agent with some of his desires as opposed to others.

These five basic points about projection show that while it can arise from other motivational states, projection generally finds its grounds in types of valuation different from the eudaimonistic value of the end to satisfy our desires or contribute to our well-being. For an
agent in projection *dedicates himself* to the pursuit of an end, rather than pursuing it simply because he finds himself *drawn* to it, for however apparently good reasons (as in S3-desires).

In this chapter, the significance of projective motivation will be confirmed by considering examples from a diverse group of authors whose work illustrates the idea of projection, though without using this term or isolating the concept. Through an analysis of contributions by Harry Frankfurt, Duns Scotus, Emmanuel Levinas, and Hannah Arendt, these and other distinctive features of projective motivation will be set in clearer relief.

### III.2. Frankfurtian Care as Projective Motivation

#### III.2.1. Caring as Challenging Oneself

The first point about projection cited above implies that when an agent ‘gives himself’ to an end of action by projecting it, the resulting sense that achieving it will be satisfying is both *derivative* (not a prior ‘pull’ on the motivation) and ‘satisfying’ in quite another sense than that of fulfilling the agent’s own well-being. Thus something often confused with a desire in an S2 or S3 sense may result from such S4-projections on ends, but the experience of wanting to achieve one’s projected goal is not the original ground of the projection; rather, it is only a byproduct of the agent’s own initial ‘push’ towards an object, which first *makes it* an end for him in the projective mode. As Frankfurt says in his essay on "The Importance of What We Care About,"

A person who cares about something is, as it were, invested in it. He identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits, depending on whether what he cares about is diminished or enhanced. Insofar as the person’s life is in whole or in part *devoted* to anything, rather than being merely a sequence of events whose themes and structures he makes no effort to fashion, it is devoted to this....Thus caring about something is not to be confused with liking it or with wanting it; nor is it the same as thinking that what is cared about has value of some kind, or that it is desirable
Like projection as I have described it, then, caring consists in an agent’s devotion to some end. This motivation precedes the conditions of satisfaction which it determines: the agent makes himself susceptible to gains or losses in happiness by committing himself to an end. In desire, it is the other way around: the perceived dependence of well-being or satisfaction on the end makes the agent desire its realization (or its non-realization).

Projection, then, as the term implies, is an essentially future-oriented kind of motivation: by placing an end out in front of himself, the agent in a sense ‘projects himself’ forward, or determines the shape of the futural possibilities towards which he will work. In the previous chapter, I argued (with reference to Frankfurt’s discussion of means and ends) that such projective motivation explains how we can first come to take up challenges that make life interesting and meaningful. Since the best sorts of human lives involve goods internal to practices that are challenging, or not easy to master, a full and meaningful life depends on projecting challenging ends and goals for their own sakes, and (importantly) not for the sake of the "good" they may be said to add to our life. Similarly, to care about something is to challenge ourselves in some way, and thus the possibility or caring depends on the same background conditions of human finitude as do practices with their own internal goods. As Margaret Watkins has insightfully argued, developing Frankfurt’s analysis,

\[\text{It is a necessary condition of things that we value highly and about which we care deeply that those things are scarce, fragile, and ephemeral. This is true of both animate and inanimate objects of caring, as well as activities about which we care.}\]

Thus, as she says, activities to which we devote ourselves would be of no interest "if

\[\text{[in the S3 sense].}\]

\[\text{Frankfurt, "The Importance of What We Care About," in The Importance of What We Care About, p.83.}\]

everyone could do everything perfectly that he/she desired with no effort or no danger or no fear of failure." The point is not that we can only care about or invest ourselves in achievements that will distinguish us comparatively from others, although the idea of distinguishing oneself is one important ground for projective motivation (among the others we have noted). Rather, the point is that the difficulty of realizing an end, or the scarcity or uniqueness of some object, provide opportunities to give ourselves challenges where none necessarily existed as an in-built part of our nature. We can make virtually any limitation the occasion for motivation, whether or not we naturally drawn to desire what is only available or possible with a struggle. We need not want to reach the summit of the mountain in any substantive sense (preferring to have a good view, or believing that the air at the top is pleasant, or judging that the activity of climbing is good for us, etc.), but the difficulty of the ascent allows us to make a project of getting to the summit.

Of course the scarcity of things we need for survival, like food and water, or the scarcity of social objects that form generic means for other ends (such as money and professional degrees) may heighten our (direct or derived) natural desire for these things. Scarcity of things we naturally need is thus a reason why they should be important to us, whether or not we regard them as important. But as projective beings, we can make other things important to us that otherwise would lack importance. Thus Frankfurt says: "It would be a serious mistake to believe that the importance of an object to someone is not fully genuine unless it is independent of his caring about the object." By caring about something, we imbue it with importance for us. As he notes, this does not mean that there are no standards for judging what is worth caring about or considering important: someone may

---

446 Frankfurt, "The Importance of What We Care About," p.93.
make a project of "avoiding stepping on the cracks in the pavement," and this is simply (as we say for worthless projects) a waste of time. There must, then, be criteria by which some cares are more worthwhile or better-considered than others: what my analysis suggests, as we have seen, is that these grounds for projection cannot themselves reduce to considerations of preference, nor even of eudaimonia.\textsuperscript{448}

III.2.2. Care and Self-Unification over Time

As the future-orientation of caring suggests, projection cannot be a simple ‘synchronic’ state of the agent in some time-slice: it must have substantial diachronic continuity in order to constitute the agent’s having a project. Projection as a motivational state is thus at least in part dispositional. In terms of the distinction with which I began this chapter, the commitment or resolve to pursue the end which projection involves is a ‘thick’ motivation because it is not equivalent to any momentary decision to act on a given motive. It is not a ‘thin’ or potentially fleeting synchronic state, but exists only as an enduring theme with importance for the agent.

In this respect, again, Frankfurt’s conception of "care" can be compared to projective motivation. Frankfurt says that caring "coincides in part with the notion of something with

---

\textsuperscript{447}Ibid.

\textsuperscript{448}It is precisely such grounds for which Frankfurt is really asking when he writes: "we need to understand better than we do ... what conditions must be satisfied if it is to be important to us to make something important to us which would not otherwise have such importance" (Ibid, p.93-94). This is equivalent to asking what the most appropriate grounds are for projective motivation (a large question, as we have seen). But note that Frankfurt’s terminology in this essay is a bit misleading, because it tends to assimilate the worthiness or value of projections with the primary/underived "importance" of their ends (as opposed to secondary of conveyed importance). This equation is misleading for two reasons: (1) primary or underivative "importance" is broad enough to include reasons to desire rather than project ends (e.g. nutrition is important for us because we need it for survival and it is scarce); and (2) the value of projecting an end may not turn solely on the values of realizing that end.
reference to which the person guides himself in what he does with his life." It is not simply an intelligible pattern that his actions exhibit over time to third-party interpreters, but the basis on which the agent herself understands the development of her course of action: this is what Frankfurt means by 'guidance.' Caring is thus "inherently prospective" and the agent who cares "necessarily considers himself as having a future." In this respect, cares are unlike desires and beliefs, which "can occur in a life which consists merely of a succession of separate moments, none of which the subject recognizes —either when it occurs or in anticipation or memory— as an element integrated with others in his own continuing history." To care is not just to have any sort of plan or project to guide one’s actions, but to have a special kind of project that expresses our intrinsic values, and the "sort of person" we are: care thus guides not just our actions but our self. These points suggest that care is an inherently projective concept.

Partly for this reason, as we will see in more detail in Chapter IV, genuine cares qualify as direct expressions of our agency and embody our personal authority; they cannot be disassociated from us in the way that any state of desire, emotion, or other psychic attitude may be. The agent is essentially ‘active’ in her cares. As Frankfurt says:

To care about something is not merely to be attracted by it, or to experience certain feelings. No one can properly be said to care about something unless, at least to some degree, he guides his conduct in accordance with the implications of his interest in it. This means paying attention to it and to what concerns it; it means making decisions, taking steps. Thus, with respect to those things whose importance to him derives from the fact that he cares about them, the person is necessarily active.

Caring is thus a paradigm volitional state in the ordinary prephilosophical sense: it embodies

---

449 Frankfurt, "The Importance of What We Care About," p.82.
450 Ibid, p.83.
452 Ibid, p.10.
the agent’s personal commitment, and cannot (in good faith) be disassociated from him. For this reason, care illustrates the notion of volition as a distinct type of motivation that binds the self together in integrity, as opposed to the thin philosophical gloss volition generally receives as the decision to act on some intention or other:

...a person can care about something only over some more or less extended period of time. It is possible to desire something, or to think it valuable, only for a moment. Desires and beliefs have no inherent persistence....If we consider that a person’s will is that by which he moves himself, then what he cares about is far more germane to the character of his will than the decisions or choices that he makes. The latter may pertain to what he intends to be his will, but not necessarily to what his will truly is.  

Similarly, the mere decision to project an end does not necessarily entail the actual substantive motivational state of projecting that end. The agent’s will in the thick sense—the motivations which constitute her identity as an agent— is thus determined by projections and cares that cannot be equated with momentary decisions.

Finally, although I have compared Frankfurt’s conception of care to my notion of projection, and in turn interpreted projection as the distinctive meaning of volition in its motivational (or ‘thick’) sense, I should acknowledge that Frankfurt does not simply equate care with volition. Instead, to care about something is a compound state for Frankfurt, which involves not only a volitional core but also the “feelings, attitudes, and interests” that normally go along with projecting a given end. We do not in this sense fully care about something (or someone) unless we also acquire the beliefs, judgments, emotions and dispositions that should come from willing its (or their) flourishing. Thus Frankfurt says that "The fact that someone cares about a certain thing is constituted by a complex set of

---

453Ibid, p.84.

454In the omitted final chapter, I consider whether there is any freedom (in the strong libertarian sense) in determining one’s projections, and how the liberty to decide in different ways fits with the concept of caring.
cognitive, affective, and volitional dispositions and states."\textsuperscript{455} We might say that one’s care represents the complete motivational \textit{mood} of a person with volitions. I have focused on the volitional or projective core of this complex mood in an effort to clarify the differences between basic types of motivational state, before attempting to consider compound states or moods in which more than one type of basic state is synthesized. This has helped to show, as promised at the outset, that will is one of the \textit{fundamental and irreducible} kinds of motivation, and not itself a composite phenomenon.

\textbf{III.2.3. Care and Love}

In "The Importance of What We Care About," Frankfurt only briefly mentions that there are many "different \\textit{ways} of caring," and that "The most notable of these are perhaps the several varieties of love."\textsuperscript{456} In a later paper, however, Frankfurt provides an analysis of "love" as a distinctively volitional kind of motivation. His account provides a particularly helpful illustration of the volitional nature of projective motivation:

As I deploy the concept, it has a very wide scope: love is a species of caring about things, and its possible objects include whatever we may care about in certain \textit{ways}...To love differs from having feelings of a certain type, such as those of powerful attraction or of intense desire or compelling delight. It is also not equivalent to or entailed by any judgment or appreciation of the inherent value of its object. To love something is quite different from considering it to be especially appealing or precious....Of course, love does ordinarily involve various strong feelings and beliefs that express, reveal, and support it. The heart of love, however, is neither affective nor cognitive. It is volitional.\textsuperscript{457}

And we might add, such love is ‘volitional’ in just the projective sense I have outlined above. For the person who loves something in Frankfurt’s sense \textit{projects} its flourishing as

\textsuperscript{455}Ibid, p.85.

\textsuperscript{456}Ibid, p.85.

her end, rather than simply responding to some prior intense admiration or fondness for it or desire to promote it. Of course, there are other motivational states (particularly S2 and S3 states) that are commonly called ‘love,’ but Frankfurt points to a kind of love that differs from these in its motivational structure. The example he gives is a parent’s "wholly active and unconditional" love for his child, which he distinguishes from intrinsically valuing the "activities and experiences" involved in caring for the child. Such a love is "unquestionably personal" rather than deriving from impartial norms and yet "utterly selfless," because "it can be satisfied completely and only by the satisfaction of interests that are altogether distinct from and independent of his own," namely, the child’s. Frankfurt connects this counterexample to Williams’s notion of motivation with the blending of ends and means in his earlier Iyyun article:

Although active self-love as such is valuable to the lover only for the sake of the benefits that it provides to his beloved, it is also true that it is valuable to him for its own sake. The loving is valuable to him for its own sake, however, precisely and only because of its utility. Serving the ends and interests of his beloved is something that he values as an end-in-itself. If he did not consider his loving devotion to be instrumentally valuable in providing benefits to his beloved, it would not be intrinsically valuable to him.

This description is likely to seem paradoxical or confused at first glance, since it might appear that Frankfurt has described a simple combination of belief and desire: I desire the other’s good, and believe that certain actions towards the other will promote that good, so I have an instrumental motivation to perform those actions. But what Frankfurt means, as I read him, is rather different. It can be summarized in three propositions: (a) the (volitional)

---

lover projects the others’s well-being as his end;\textsuperscript{461} (b) he also projects the end of providing the other with his loving devotion; and (c) the ground for this latter projection is his belief that his devotion to the other will promote the other’s well-being. In this case, then, one projection grounds another, but this relation is not like the ordinary end-means relation. It would be strange to say that our agent desires to show loving devotion to the other because this is a means to serving the other’s interests: such loving activity cannot be an object of desire, i.e. something apprehended as lacking in the agent’s eudaimonia (unless his need to be of use to the other is more pathological). Instead, in the normal case, loving devotion is an activity that, in its very nature, involves motivation of the projective kind: it must be projected as an end, not a mere means, which implies that the agent values the activity of pursuing the other’s good intrinsically. Yet his reason for projecting this activity of devotion and thus valuing it intrinsically is its instrumental relation to the other’s good. This interpretation illustrates why Frankfurt’s attempt to distinguish his model from the earlier Aristotelian model of motivation can only be made fully intelligible in light of the structure of projection.

So understood, Frankfurt’s point might be buttressed by Elizabeth Anderson’s remarks about what she calls the "attitude" of love. These remarks are made in the context of an argument against the adequacy of a consequentialist principle, P, which says "Act so as to promote the state of affairs: that my actions adequately express my rational attitudes

\textsuperscript{461}Frankfurt actually says that "The active lover values the ends and interests of his beloved for their own sake," rather than referring to their well-being or happiness (note 4, p.439). But since a good father, friend, or concerned person may care for the other by opposing the other’s subjective ends and interests when these will not really lead to the other’s eudaimonia, it seems better to revise the formula as I do in proposition (a). This only makes clearer the point that the other’s well-being is not the lover’s own eudaimonia, and therefore (like certain forms of altruistic motivation) it presents a problem for anyone who believes only in teleological motivation and not projection.
towards persons.” Anderson argues that this principle, in asking us to maximize a certain state of affairs, implies the wrong motives for action:

In acting on P, I take my own desire to be a loving person as my reason for action. In taking this to be my reason for action, I do not act out of love for my husband....What makes sense of my taking my husband’s needs as my reasons for action is that I love him, not that in so taking them I make myself a loving person. Anderson has imagined here another possible motive for acts of loving devotion, which would replace (c) in our analysis of Frankfurt’s example above: namely, the desire to validate a certain ‘self-image’ of herself as loving, which is measured by how much loving activity she exhibits. This is state—namely, a desire to see herself as loving— is one possible motivation that would have the consequence-seeking structure recommended by P. But action on this motive is not genuinely loving, because it does not take her satisfying her husband’s needs as intrinsically valuable, and project the end of exhibiting loving activity only on the basis of its usefulness to this primary end, as Frankfurt says. Notice that Frankfurt’s volitional lover would be willing to sacrifice his self-image as ‘the person lovingly devoted to person A’ if he saw that his loving devotion could not benefit A, whereas the putative lover acting on Anderson’s instance of P would not. The projective structure of volitional love implies that the true lover would sometimes be motivated to sacrifice her loving activity out of love itself. A desire, by contrast, can never be motivated to abandon pursuit of its end for the very sake of this end, because to abandon the pursuit means that the agent himself will not be the beneficiary of any realization of this end. An end that can only be realized for another by means of my abandoning its pursuit, is one I cannot desire, if to desire is to want the requitement of a lack, as I have argued. Such an end

---

462 Anderson, op. cit., p.543. This principle was suggested by Sturgeon as an alternative formulation of Anderson’s own neo-Kantian principle of respect for persons.

463 Anderson, p.545.
can only be projected. To put this in plainer English, I cannot sacrifice my enjoyment of some object or goal for another out of desire for this very object or goal. Such a sacrifice could be desire-motivated for me only if it somehow helped me realize some other end. But volitional love includes the possibility that it may sometimes motivate the lover to abandon pursuit of his end, not because this sacrifice is satisfying to him in any other way, but precisely for the sake of the end abandoned, which desire can never do.

### III.3. Williams’s Neo-Humean ‘Internalism’

One reasonable reaction to the foregoing analysis of will as projective motivation might be that it is simply the familiar story of anti-Humean "externalism" in a different guise. I claim that from certain evaluative beliefs, which I call "cares," an agent may acquire new motives without going through an intermediate step of creating a desire. One kind of these beliefs are moral judgments, which entail a motivation to form intentions, independently of intervening desires: this is just one version of "externalism."

I make two points in response. First, "cares" are not simply dispassionate evaluative beliefs that something is important or worthy to care about; they already involve what I call first-personal appropriation, or in Frankfurt’s terms, "identification." Second, there are not only a confusing variety of different senses of ‘externalism’ these days, but also several possible versions of each. Sverdlik and others make it part of ‘internalism’ that certain beliefs entail motivations and thus intentions$^{464}$ (although, as we saw, he does not build into the definition that there that the causation is direct, without intermediary desires). As David Velleman explains, according to Bernard Williams,

---

"internal" reasons are those which count as reasons for someone only by virtue of his antecedent inclinations; "external" reasons are those which count as reasons for someone [to act] independently of his [prior] inclinations. The Humean conclusion implies that all reasons are internal, in this sense...⁴⁶⁵

This formulation leaves room for several types of an anti-Humean "externalism." One version, which can seem deceivingly similar to internalism, would hold that at least moral reasons (and possibly other types of practical reasons as well) always entail a motivation to act, but not on the basis of any motivation in the agent’s internal set prior to recognizing the external reason. Call this strong moral externalism: on this theory, recognizing moral reasons formulated in impartial, third-personal terms, must always be the source of a new motive in the agent’s internal set.

A weaker version of moral externalism —and one that leaves more room for freedom— would hold that certain beliefs can lead to motivations that do not arise from pre-existing desires in the agent’s ‘internal set.’ A specific variant of this —which I’ll call ‘projective moral externalism’ (PME)— would hold that certain beliefs can lead to (but do not entail) related new motivations, and thus intentions and actions, without passing through desires at all (and thus a fortiori without dependence on desires in the agent’s pre-existing ‘internal set’). PME is the version of moral externalism most consonant with my analysis of the will in terms of projective motivation. And PME allows for the possibility that first-personal motivation may be formed on rational grounds that abstract from the desires in an agent’s internal set. Thus PME poses a serious challenge to Williams’s well-know argument that Hume is basically right that "all external reason statements are false" because in the cases that would disprove Hume, "ex hypothesi, there is no motivation for the agent to

deliberate from, to reach this new motivation.” My account of projective motivation at least begins to indicate how it is possible for an agent to give herself new motives on the basis of grounds or considerations that are presently external for her in the sense that her recognition of their practical relevance is not derived from their apparent connection to how she could better realize the set of ends towards which she is already motivated.

At this point, the Humean will undoubtedly ask why she gives herself a motive. However, this question is ambiguous. If it means to ask what her motive was for giving herself a motive, then the question implies the false presupposition that every motive must either have a prior motive, or simply be brute (and thus not undertaken). Volitional motivation is precisely the kind for which neither of these alternatives is true. If the question means to ask instead what her reason for motivating herself was, then it will be answered by citing the grounds that were relevant for her, but according to PME, her judgments about different ends she might pursue, and the pursuit of them itself, does not entail her subsequent projective motivation. Thus the theory of projective motivation can bridge the gap between reasons to act that are purely cognitive or motivationally impotent for the agent, and the brute desires or appetites that Humeans believe give cognitive considerations their motivational potential, because it is a middle soul state, which is neither merely cognitive nor merely appetitive: it is the will through which the agent makes new motivation out of the raw material of cognitive resources that do not automatically motivate on their own. In this perspective, we can see that it is precisely because will in the sense of a middle-soul state is missing from contemporary moral psychology that the debate is dichotomized between rationalist Humeans and rationalist neo-Kantians. If this stalemate is to be broken, it depends on recognizing the reality of projective motivational states that are neither derived

---

from non-self-determined appetites nor entailed by external reasons.

This also explains why PME differs from Bernard Williams’s construal of the what an anti-Humean position would have to be. Succumbing to the dichotomy, Williams defines anti-Humean internalism in terms of an implausibly strong rationalism about motivation, which would hold that intentions and action can result directly from rational beliefs without any first-personal motivation at all. The notion that this implausible view is the only serious alternative to the Humean position is a straw man, as the alternative represented by PME reveals.

Surprisingly, though, there remain questions about how substantially Williams’s own conception of motivation really differs from the projective conception I have laid out. For if the Humean not only can allow his ‘internal set’ to be expanded by rational deliberation, as Williams claims, but could also allow it to be expanded with motivations that are not desires, the Humean thesis would no longer be at all distinctive or interesting.

III.4. Duns Scotus on Free Will

So far, I have traced the suggestion that something like projective motivation is possible back only as far as Kant. But the idea I am seeking to develop has far deeper roots, probably going back as far as the Stoics and Augustine. Although I will not try to trace it back that far, it will be helpful to see that, in contrast to Aquinas, the leading author in medieval Franciscan philosophy analyzes the will in terms that require projective motivation as I have described it.

Dun Scotus’s conception of the will is of course an extremely complex matter, but I will only touch on a few of its most important points. First, in his famous Oxford Commentary, Scotus explicitly rejects the notion that "will" (voluntas) properly so called is
moved by motives prior to it. As I have (Chap. II §3.3), he attributes this view to Aristotle: "In Book III of the De Anima Aristotle lays down an order of what moves and what is moved, thus: the desirable object is an unmoved mover; the appetite is a moved mover, moved, that is, by the desirable object..." Citing Augustine, Scotus argues instead that "it is in the power of our will to have negative and positive attitudes to a single object. Therefore these cannot be produced by an agent naturally; and therefore not by the object, which is a natural agent." We can view this as an argument that a faculty involving libertarian freedom with respect to the same intentional object of action could only be motivated in a projective fashion. My prior arguments for projective motivation have not presupposed libertarian freedom in this fashion, but Scotus has other arguments as well.

Second, in reply to the type of position Aquinas maintains, Scotus agrees that the will controls the intellect by being "able to move the intellect to the consideration of this or that," but argues that this itself cannot be determined by the natural and involuntary orientation of the will toward the Good, or else "it will no longer be in the power of the will so to command the intellect concerning the consideration of this or that any more than was the first act" of desiring the Good. Scotus concludes that "nothing other than the will is the total cause of volition in the will;" he regards the will as an indeterministic cause of action, and unlike Aquinas, does not rest this indeterminacy in the intellect’s power of attention: "There is a different form of potency which is still undetermined after the object...

---


468Ibid, Reply to the First Position (a), p.36.

469Ibid, Reply to the First Position (c), p.37. He adds that it explains nothing to say that the phantasm presented to the intellect has negative and positive aspects, either of which can move the will depending on which it focuses on, because if these two external motives are equal in force, the will is paralyzed, and if not then it is determined to go one way rather than the other (Reply to the First Position (e)).

has been presented to it, which is perfect, and not diminished. The will is of this kind...

There is evidence that Scotus changed his teachings on this matter somewhat in later lectures given after his original move from Oxford to Paris, which were not originally part of his Oxford Commentary. In these lectures (as recorded by reliable students) Scotus no longer holds that the object of the will is merely a causa sine qua non, but teaches instead that the object is a partial cause of the will’s movement towards its end. The free choice and the object are independent causes of the action in the sense that neither depends on the other for its role in influencing the outcome, but freedom is still "the principal cause." Yet even with this modification, Scotus’s position remains original. Alan Wolter explains why in his discussion of Scotus’s fifteen questions on Book IX of Aristotle’s Metaphysics, where the Master challenges Aristotle’s principle that "Whatever is moved is moved by another." As Wolter summarizes:

Among other instances of "self-movement," Scotus singles out the human will’s ability to determine itself. As an active potency, this will is formally distinct from, but really identical with, the soul substance, and is either the exclusive or at least the principal efficient cause of its own volition. The volition in turn is an immanent action that falls under the Aristotelian category of quality...[so] one can correctly say that the soul "moves itself" from a state of indeterminacy to a positive state or decision.

Wolter suggests that Scotus’s notion of the "superabundant sufficiency" of the will to

---

471Ibid, "The Principal Arguments are Resolved" (d), p.39.


474Ibid, pp.110-111.

475In the Wadding Reprint of Scotus’s Opera Omnia, tome VII, this work is entitled: Quaestiones subtilissimae super libros Metaphysicæ Aristotelis.

476Actually Scotus, being overly charitable, denies that Aristotle really intended any such principle.

determine itself refers to a state "more suitably called ‘positive indeterminacy’ rather than ‘negative indeterminacy,’" since it is not arbitrarily determined but determines itself in the light of considerations presented by the intellect that nevertheless by themselves underdetermine it. 478

Scotus still follows older Aristotelian definitions of the will as "rational appetite." 479 He admits the existence of motivating desires and appetites that are simply natural to the human species, the principal of which is will’s necessary and thus passive appetite for "happiness" in general, in which the "infinite good, as man’s ultimate end, occupies the first place." 480 But as a free power, the will can pursue an evil end (thus making it into "an apparent good") because

it can choose any being as the object of its desire and make it an ultimate end, were it not for any reason other than its own pleasure in abusing freedom and performing an evil act. Thus a rational creature, with full knowledge of what it is doing, can hate God and find satisfaction in such a hatred, not because God is hateful or because the hatred of God is not something evil, but because of the pleasure that even such a hatred can bring to a rational creature in the form of an apparent good. 481

This seems very close to the notion of projective motivation. Scotus keeps the notion that formally, what the will seeks must always be apprehended under the aspect of an apparent good, but he implies that this is not because the will’s free choice itself is motivated or drawn by its natural tendency to happiness, but rather because whatever end the will gives itself thereby becomes apprehended as an apparent good for it. In other words, the satisfaction of realizing the apparent good is a noncomprehensive satisfaction in the sense

478 Or so I will argue in describing a constrained version of libertarian freedom in the projected additional chapter on free will.
479 Bonansea, p.86.
480 Ibid, p.87.
481 Ibid, p.89, referring to the Ordinatio [Book One of the Opus Oxoniense], I, d.1, nn.16-17; II, pp.10-11; and Opus Oxoniense, II, d.43, qu.2, n.2; XIII, pp.493b-494a.
defined by Point (1) concerning projective motivation at the start of this chapter. As Bonansea says, this puts Scotus in opposition to Aquinas’s theory that "no specific distinction obtains between the natural appetite [for the good] and deliberate volition, since the latter is for him merely an application to a particular good or the original movement of the will towards good in general." As Bonansea explains it, Scotus’s main counterargument rests on the point, which Aquinas concedes, that in its free acts, the will can divert the intellect from particular good or prevent the agent from knowing them; then "since the will is free in directly the intellect towards its final end, or happiness in general, it is also free in its tendency towards happiness as such." Similarly, he objects against Aquinas’s theory that, if in its free or "elicited" acts, "the will tends of necessity towards happiness as its final end, and it is within its power to apply intellect to the consideration of that end, it will do so at all times and necessarily." But we know from experience that it does not.

As Wolter explains, for Scotus, this means that not only is every elicited act of the will "free in the sense of its being an act of self-determination elicited contingently and not deterministically," but also that is at liberty for different possible motives. To explain this, following Anselm, Scotus argues that in addition to the natural appetite for happiness or the agent’s eudaimonia, the will has an inclination towards "good" in another less personal sense, so that it has two different highest possible motives. But the latter is a motivation on
which the will acts only in its free elicitation of acts.\footnote{Ibid, \textit{Ordinatio}, IV, suppl. d.49, qq.9-10, Art.2, p.189.} As Wolter explains:

...there is a prior sort of "liberty" or "innate freedom" possessed by the will, a liberty that frees it from the need to seek self-perfection as its primary goal, or as a supreme value. It consists in free will’s congenital inclination towards the good in accord with its intrinsic worth or value rather than in terms of how it may perfect self or nature. Anselm calls this higher inclination the will’s "affection for justice," or \textit{affectio iustitiae}. In virtue of this...the will is able to love God for his own sake as a supreme value. But it also inclines the will to love other lesser goods...in terms of their intrinsic worth rather than in terms of how they perfect one’s individual person or nature.\footnote{Ibid, Wolter’s Introduction, pp.39-40.}

Scotus thus seems to have grasped that being \textit{motivated to pursue} certain ends for their own sakes can be \textit{good in itself} in a different sense than the happiness that realizing the end would bring the agent; this "justice" can give the will grounds for a non-teleological sort of motivation. Note how similar this description of the second inclination sounds to the notion of moral grounds as a form of projective motivation (II.7.1): although Scotus and Anselm may not conceive the criteria of "rightness" or \textit{rectitudo} in terms of a formal universalizability test, they also explicitly do not conceive it in eudaimonistic terms either.\footnote{On this topic, also see John Boler, "Transcending the Natural: Duns Scotus on the Two Affections of the Will," in \textit{Duns Scotus}, ed. Alan Wolter, \textit{American Catholic Philosophical Quarterly}, Vol. 67.1 (Winter 1993): 109-126, esp. p.122.}

This parallel makes more sense once we realize that although Scotus’s two highest innate potentialities in the will are called "inclinations" and described as "the affection for justice and the affection for the advantageous,"\footnote{\textit{Duns Scotus on the Will and Morality}, ed. Wolter, \textit{Ordinatio} III, suppl. d.26, p.179.} they are not really both \textit{appetites}, since they differ dramatically in structure. The appetite for advantage or happiness is "simply the inclination the will has towards its own perfection, just as in the case of other things that lack free appetite," and is the tendency towards its final cause that any "nature" must have to
be a nature.\textsuperscript{490} Scotus also clearly understands the natural appetite for happiness or advantage according to the lack-model of desire: he describes it as "the inclination the potency has to tend towards its proper perfection...It is imperfect unless it possesses that perfection to which this tendency inclines this power." By contrast, when the will "tends freely and actively to elicit an act," it is active rather than passive, since it posits the tendency towards the end rather than merely being a tendency caused by the need to have an object to realize perfection.\textsuperscript{491} Thus, in arguing that the agent’s own happiness is not "the rationale behind all willing," Scotus sums up the two types of motive as follows: "There is a twofold appetite or ‘will:’ one, namely, that is natural [i.e. necessary], another that is free."\textsuperscript{492} The difference is that the latter can choose its own final end. If this is right, then the affectio iustitiae is not really a form of desire at all, and Scotus’s two basic types of motivation seem to line up well with (a) the general desire for happiness as the ground of S3-desires and (b) projective motivation (as least in a less formalist version of its deontic form).

This leads to another fundamental novelty in Scotus’s conception, which connects to Frankfurt’s discussion of love. Scotus holds that agape, as love of God and neighbor for their own sakes, is a possibility built into the will rather than one requiring an infusion of grace:\textsuperscript{493}

The affection for justice is nobler than the affection for the advantageous, understanding by "justice" not only acquired or infused justice, but also innate justice, which is the will’s congenital liberty by reason of which it is able to will some good not oriented to self. According to the affection for what is advantageous, however, nothing can be willed save with reference to self. ...To love something in itself [or for its own sake] is more an act of giving or sharing and is a

\textsuperscript{490}Ibid, \textit{Ordinatio}, IV, suppl. d.49, qq.9-10, Art.1, p.185.
\textsuperscript{491}Ibid, \textit{Ordinatio} III, d.17, p.183.
\textsuperscript{492}Ibid, \textit{Ordinatio} IV, suppl. d.49, qq.9-10, p.183.
\textsuperscript{493}Although Scotus does distinguish between "the will itself in a purely natural state," including its freedom and two basic inclinations, and "the same will as informed by gifts of grace": Ibid, \textit{Ordinatio} IV, suppl. d.49, qq.9-10, p.183.
Thus what I called "pure motivation," or the ability to seek an end irrespective of the value to the agent of realizing it, or even of seeking it, is built into the will for Scotus. This implies that "it is not necessary that the will seek whatever it seeks because of its ultimate end as a source of happiness" (although this does not for him also imply that we can ever will misery or evil per se). Pure love as Scotus conceives it is thus a clear example of projective motivation. Conscience of synderesis, as its ground, is in the intellect and does not necessarily move the will. As Wolter summarizes, "Scotus, in contrast to many of his contemporaries, believed that man by his natural powers could love God above self or any other created good, and that this capacity was not something he possessed only in virtue of some special grace or the infused theological virtue of charity." As we will see in the next section, Levinas holds something similar regarding our capacity to care for others.

III.5. Levinas on Superabundant Will and Volitional Generosity

Moving back to the 20th century, Emmanuel Levinas’s well-known ethical
phenomenology depends on the view that something similar to projective motivation is possible for us, although his account of it differs from my own in one respect, as we will see.

Levinas’s philosophy is built on his argument that our encounter with the other person as a being to whom we owe an absolute responsibility, which he refers to as being called by the other’s "Face," has absolute primacy in the structure of selfhood. It is in this encounter with ethical alterity—the Otherness of the other person—that we first encounter reality outside our consciousness at all. Yet this is not something added to a self-knowing consciousness that is already self-standing; rather, it is only by being appropriated into moral responsibility for others that the I arises as a self-aware center of agency at all. It is helpful to think of this as an inversion of Sartre’s quasi-Hegelian conception of the I as threatened in its solitary self-absorption by the existence of the other’s look. Levinas is specifically opposing the philosophical tradition in which "conflicts between the [self-]same and the other are resolved by theory whereby the other is reduced to the same—or, concretely, by the community of the state..." Instead, the Other is originally an ethical appeal that cannot be assimilated back into the self or mediated in a higher synthesis: in Fichtean terms, it is the Anstoss through which the I comes to have the reflexivity necessary for selfhood and agency to begin with.

The "ethics" arising from this self-originating relation to the other is conceived not as a limitation on freedom of the will, but as freedom’s ground of possibility outside itself (since the I does not ground itself). As Levinas says, "the other absolutely other—the Other—does not limit the freedom of the same; calling it to responsibility, it founds it and

\[\text{\footnotesize 499}\] See Jean-Paul Sartre, \textit{Being and Nothingness}, Part Three: ch.1 §4 ("The Look") and ch.3 ("Concrete Relations with Others"), esp. p.482 (the comparison to Hegel’s Master-Slave dialectic).

As in Kant’s moral philosophy, the agent’s free will is possible only as already constrained by an absolute moral responsibility it did not choose. Merold Westphal’s summary of this point brings out its basic similarity to the duty of neighbor-love in the Judeo-Christian tradition:

He [Levinas] is claiming that the face of the neighbor confronts us not as a contractual proposal to be negotiated, but as an unconditional obligation. It is unconditional in that its validity depends in no way either upon our agreeing to accept it or in the Other’s doing something to evoke or merit our compliance.

Levinas describes this originary relation of responsibility for the other person, which is awakened by their "face," as a "metaphysical desire for the absolutely other." Levinas’s choice of terms here is singularly unfortunate for my purposes, but we will see that what he means by "metaphysical desire" is precisely a non-teleological type of motivation. The "desire" Levinas has in mind contrasts with Plato’s eros-model since it is not motivated by any want that reaches towards objects to satisfy itself. It is specifically contrasted with cognitive and appetitive relations to realities on which I can "feed...as though I had simply been lacking them. Their alterity is thereby reabsorbed into my own identity as a thinker or possessor. The metaphysical desire tends towards something else entirely, towards the absolutely other," or towards something I do not already anticipate. Ethico-metaphysical Desire is meant to contrast with all teleological forms of causality, intentionality, or motivation, which he considers violent in a quasi-Kantian sense: "the fabrication of a thing,

---

502Kant’s emphasis on the moral law as the faktum of pure reason through which alone we know freedom and his emphasis on the primacy of practical reason is passed along through Herman Cohen’s philosophy to Martin Buber and thence on to Levinas and Jean-Luc Nancy. Thus Levinas compares the transcendent "infinity" of the Face whose otherness I cannot contain to the infinity of ideas in Kant’s thought (Totality and Infinity, p.196).
504Levinas, Totality and Infinity, p.196.
505Ibid, p.33.
the satisfaction of a need, the desire and even the knowledge of an object" are violent in the sense that they use what is alterior as a means for self-satisfaction. Instead, as he explains in an early essay,

In the "Canticle of the Columns," Valéry speaks of a "faultless desire." He is doubtless referring to Plato who, in his analysis of pure pleasures, discovered an aspiration that is conditioned by no prior lack. We are taking up this term desire; to a subject turned to itself...we are opposing the desire for the other which proceeds from a being already gratified and in this sense independent, which does not desire for itself. It is the need of him who no longer has needs.... The desire for the other, sociality, is born in a being that lacks nothing, or, more exactly, it is born over and beyond all that can be lacking or that can satisfy him.... The movement towards the other, instead of completing me and contenting me, implicates me in a conjuncture which in a way did not concern me —what was I looking for here?

Here Levinas clearly contrasts our original ethical motivation to care for the other with all forms of "desire" in the substantive sense, i.e. teleological motivation in which the agent wants an end to restore an equilibrium by making up for a perceived lack in the agent’s well-being (Chap. II §3.2 & §5.2). Instead, he thinks of metaphysical Desire in specifically neo-Platonic terms as a kind of overflowing of an agent that, in this motivation at least, needs nothing for itself, but goes freely out of itself towards an end it was not drawn to realize by any prior apprehension of an imperfection which the end would satisfy. In another essay, Levinas explains that he talks about the infinity of metaphysical desire "to mark the propulsion, the inflation, of this beyond" since it is opposed to the affectivity of love and the indigence of need. Outside of the hunger one satisfies, the thirst one quenches and the senses one allays, exists the other, absolutely other, desired beyond these satisfactions...This desire is unquenchable, not because it answers to an infinite hunger, but because it does not call for food.


__507__Levinas here cites Plato’s *Philebus* 50e ff and *Republic* 584b ff.

This desire without satisfaction hence takes cognizance of the alterity of the other.\(^{509}\)

By metaphysical desire, then, Levinas means something like a *projection* in which we become devoted to an end for some other reason that either the satisfaction it promises or its instrumental value in facilitating such satisfaction. This motivation is unquenchable because it can acquire ever-new ends, since their acquisition does not arise from a sense of lack wanting fulfillment.

Although metaphysical desire is more like a *duty to care*, in response to which we can take on the other’s concerns as our own, Levinas describes this sense of obligation as if it were *itself* a form of neighbor-love. After explaining why the "Infinite" is unlike all intentional contents of thought, since it "designates a relation with a being that maintains its total exteriority with respect to him that thinks it," Levinas says that a relation can nevertheless be maintained with what thus exceeds representation through

Desire—not a Desire that the possession of the Desirable slakes, but the Desire for the Infinite which the desirable arouses rather than satisfies. A Desire perfectly disinterested—goodness. ...For the presence before a face, my orientation toward the Other, can lose the avidity proper to the gaze only by turning into generosity...\(^{510}\)

This formulation parallels the oldest definition of agape as disinterested love, unselfish generosity as opposed to investment for a return.\(^{511}\) As Levinas puts it later, the "proximity" or relation of the self for-the-other "is a for of total gratuity, breaking with interest."\(^{512}\) He even refers to the responsibility that requires substitution of oneself for another—the

\(^{509}\) Levinas, "Philosophy and the Idea of Infinity," in *Collected Philosophical Papers*, p.56. Note that by "love" here Levinas is referring to erotic love.

\(^{510}\) Ibid, p.50.

\(^{511}\) Compare this to Levinas’s remark in his essay, "God and Philosophy," reprinted in *Collected Philosophical Papers*, p.163: "But this desire is of another order than the desire involved in hedonist or eudaimonist affectivity and activity, where the desirable is invested, reached, and identified as an objective of need....It is a desire that is beyond satisfaction, and, unlike a need, does not identify a term or an end. This endless desire for what is beyond being is dis-interestedness, transcendence — desire for the Good."

sacrifice of subjectivity— as "love without eros." Such a responsibility "to answer for the other is, perhaps, the harsh name of love." Similarly, near the end of his famous essay on "Ethics as First Philosophy," Levinas describes responsibility as "love without concupiscence." He equates this with a "mauvaise conscience" not arising from any action or intention, an originary guilt which gives us a "capacity to fear injustice more than death, to prefer to suffer than to commit injustice." Levinas also describes this revelational relation as "the binding separation known by the well-worn name of love." In other words, the face-to-face relation has the same structure which Derrida finds in an ideal (or, for him, "impossible") gift beyond all economic investment or expectation of a return.

Levinas’s interest in the non-teleological structure of agape is directly related to his opposition to Husserlian phenomenology, in which the content thought is the correlate of the mode of consciousness or the intentional attitude directed towards it. His hostility to this model stems from the analogy between its structure and the structure of ordinary appetitive motivation as lack. Just as the object desired is proleptically anticipated by the desire for it, since it is apprehended in terms of its value for the agent’s well-being or satisfaction, Levinas believes that in Husserl’s model, all intentional contents are conformed to the conscious act of the agent (as in Plato’s maieutic theory of knowledge, the object known is anticipated in the agent’s trace memory of it). In such conceptions, Levinas believes, we never find in the world anything but reflections or externalizations of what was already within us: that is, we never encounter genuine otherness without constructing it in our own

---


In opposition, he emphasizes that in metaphysical desire, the Other thus does not become our end by first being cognitively appropriated as the correlate of some gap in our happiness; instead, it retains a meaning that is not constructed by the intentionality of our attitudes towards it at all: Original responsibility "in nowise resembles the intentional relation which in knowledge attaches us to the object," at least if this is understood as an object determined by the mode in which the subject intends it, as in classical phenomenologies. As in agape, the agent’s orientation to the Other is not an intentional relation in the classical sense: in both cases, the agent moves towards an end radically exterior to her self. This similarity motivates Levinas’s suggestion that the essential dependence of personal interiority on responsibility for others and for the human "world" in general is a "fundamental non-narcissism.". As in agape, the Other towards which the agent is oriented is not an "end" in the classical sense of a correlate that fulfills some deficit in the agent, or addresses some lack which would motivate a self-interested concern to realize that end. In this sense, the relation in both cases is non-teleological or "dis-interested," as Levinas says: the end is not "grasped," or does not become an object or goal

---

518This position developed in particular through Levinas’s reaction to the account of our experience of the Other in Husserl’s fifth Cartesian Meditation, and its descendent in Sartre’s discussion of “The Look” in Being and Nothingness.


521Compare this to Levinas’s commentary on a Talmudic passage discussing how we can have any relation to God as the Infinite or unnameable En-Sof. We cannot say that we "think" the Absolute God, since: "Does not this word conjure up, if not vision, then at least aim, which in its way posits another end or sets it as its target? The text we have just quoted suggests a beginning that does not move towards an end, but traces, as it were, a relation without correlate" (Levinas, Beyond the Verse, p.165). This is typical of his hyperbolic style: to emphasize how different this relation is from those of classical intentionality, Levinas suggests that it cannot be described as a relation to an end or content of intentionality at all.

522Levinas, Ethics and Infinity, p.100. Levinas splits the word this way to suggest that such non-teleological (continued...)

244
purely through its relation to the agent’s intending; rather, the agent leaves relation-to-herself behind in going out towards this goal. Thus "the vision of the face is not an experience" or absorbing of objects into the self, "but a moving out of oneself, a contact with another being....The infinite is given only to the moral view: it is not known but is in society with us."  

Similarly, in commenting on Franz Rosenzweig, Levinas describes "revelation" as an "entering-into-relation completely different than the one that corresponds to a synthesis"; since such a relation is not teleological, it links its terms non-additively "in a connection for which language or sociality or love is the original metaphor."

The connection between Levinas’s descriptions and my account of projective motivation has already been indicated. Levinas’s Neo-Platonic metaphor of superabundance is perfect for projection, because such motivation is an expression of the creativity of the agent, who gives himself an end that is not proleptically anticipated as needed for his completion. In this sense, it is a motivation that overflows onto new objects and goals, rather than returning back to his own needs and wants (however enlightened) as its basis. Similarly, Levinasian metaphysical desire is unrelated to any vested interest of the agent: it "implies the disinterestedness of goodness." Thus Levinas connects this with Plato’s notion of a Good beyond Being, which he understands as an unappropriable object: metaphysical desire "desires beyond everything that can simply complete it. It is like

---

522(...)continued) orientation involves a relation unlike those terminating in "being" (esti). For example, the ethical "face" of the other cannot be grasped or "become a content" in the classical sense: "It is in this that the signification of the face makes its escape from being, as a correlate of knowing" (Ibid, p.87). This is the sense of "being" in which the Face is "beyond being."

523Levinas, *Difficult Freedom*, p.10. Note again the similarity to Kant’s notion of the primacy of a practical reason the core of which is given in an absolute faktum.

524Levinas, *In the Time of Nations*, p.158.

525Levinas, *Totality and Infinity*, p.35.
goodness—the Desired does not fulfill it, but deepens it.\footnote{Ibid, p.34.} In other words, this is a case where the ground of the motivation is the inherent value of pursuing the end rather than the eudaimonistic value of realizing it—a case fitting the paradigm of projection described earlier (Chap. II §6).

Levinas’s discussion of the will confirms these connections. It is in willing that we find the individual’s most intimate reflexive relation with himself, his capacity to determine the motives on which he acts:

When the will triumphs over its passions, it manifests itself not only as the strongest passion [contra Hobbes] but as above all passion, determining itself by itself, inviolable. But when it has succumbed, it reveals itself to be exposed to influences.\footnote{Levinas, \textit{Totality and Infinity}, p.257.}

The will can transcend all appetitive motivation. But, in keeping with the existentialist emphasis on the factual side of selfhood, Levinas emphasizes that the will characteristic of persons is always exposed to violence from others, such as threats, pressure, coercion, manipulation, hatred which imposes suffering that reifies the will, and even murder.\footnote{See Levinas, \textit{Totality and Infinity}, pp.226-247.} Our volitional activity is dependent on relations to others, who take up the results of our work, and can by their own activity give a significance to our efforts that we did not intend. As Levinas says,

...this way a will plays in history a role it has not willed marks the limits of interiority: the will finds itself caught up in events that will appear only to the historian.\footnote{Ibid, pp.227-8.}

The whole being of willing is hence not enacted within oneself. The capacity of the independent I does not contain its own being.
As John Llewelyn says, "Levinas located will in the lived body." Yet its freedom is restored to the will, paradoxically, by its being called beyond itself in absolute duty, in "will as Desire and Goodness limited by nothing." It is the presence of moral grounds it does not choose that allows freedom to be sovereign for the individual, since without this metaphysical dependence, freedom is arbitrary. Yet neither does this conception avoid arbitrariness by reducing the will to reason; rather, employing the neo-Platonic metaphor again, Levinas says that will requires "the idea of creation which, in God, exceeds a being eternally satisfied with itself." Levinas conceives such creative motivation as neither purely cognitive nor affective, but rather as an irreducible form that stands in between these two —like the ‘middle part of the soul.’

But someone will object that there is a fundamental difference between my account and Levinas’s, since he conceives metaphysical "desire" of being obligated unconditionally to the other as involuntary, a duty that is the very ground of possibility of agency, whereas projection was described as voluntarily motivating oneself. But note that I also described projective motivation as always requiring grounds; also note that Levinas believes the agent’s response to this involuntary call to the ‘Face’ can be purely motivated, precisely because of the nature of its ground: "The will is free to assume this responsibility in whatever sense it likes; it is not free to refuse this responsibility itself [or] to ignore the

531Ibid, p.239.
532Levinas, Totality and Infinity, p.303.
533Ibid, p.218. He adds, in Kierkegaardian fashion, and against Sartre, that "If the subjectivity were but a deficient mode of being, the distinguishing between will and reason would indeed result in conceiving the will as arbitrary, as a pure and simple negation of an embryonic or virtual reason dormant in an I...If on the contrary, the subjectivity is fixed as a separated being in relation with an other absolutely other,...then the will is distinguished fundamentally from the intelligible, which it must not comprehend, and into which it must not disappear, for the intelligibility of the intelligible resides precisely in ethical behavior, that is, in the responsibility to which it invites the will" (p.218).
meaningful world into which the face of the Other has introduced it.” Thus metaphysical desire plays the same role in Levinas’s account as the ‘inclination to justice’ in Scotus’s account: this inclination, which is itself involuntary, is the basis for the will’s freedom to determine motivations not given to it by the intellect’s apprehension of objects as ‘good for itself.’ And just like Scotus, who treats this ground as if it were itself a motive, but characterizes this motive in non-teleological terms, Levinas transfers the structure of the projective motivations it makes possible (e.g. agape) back onto the sense of absolute ethical beholdennes that grounds them. The only formal difference between these accounts and mine is that I do not treat grounds of projections as themselves constituting motives of this kind.

III.6. Arendt’s Distinction Between Circular and Linear Motivation

It will be helpful to compare the projective account of will outlined in Chapter II with the views of one other recent continental philosopher who has made original contributions on this topic. In volume II of The Life of the Mind, Hannah Arendt not only retraces the history of conceptions of willing from Aristotle through Scotus and on to Nietzsche and Heidegger (thus covering most of the same topics I have discussed in my Introduction, Chapter I, and Chapter III), but also proposes some novel points of her own, on which I will focus here.

Arendt begins by with the existential relation of will to time: "the Will, if it exists at all...is obviously our mental organ for the future as memory is our mental organ for the past. The moment we turn our mind to the future, we are no longer concerned with ‘objects’ but

---

with *projects*... By this Arendt means projects in the literal sense of plans and directives, e.g. those spelled out in a "Last Will and Testament." But Arendt follows Augustine and Scotus in conceiving the will as essentially free in a way that (as we will see) requires the possibility of projective motivation:

A will that is not free is a contradiction in terms—unless one understands the faculty of volition as a mere auxiliary executive organ for whatever either desires or reason has proposed.\(^{537}\)

By contrast, Arendt conceives the will as a motivational faculty of the *middle-soul*, not determined by appetite or evaluative judgments. Moreover, she rejects recognizes that volitional motivation cannot have the kind of teleological structure that Aristotle invokes to explain "making or fabrication—*poiein*, as distinct from *prattein*, acting or praxis." The problem with Aristotle’s account is that the model in which "everything that appears grows out of something that contains the finished product potentially" makes time into a circle, denying genuine novelty:

The view that everything real must be preceded by a potentiality as one of its causes implicitly denies the future as an authentic tense: the future is nothing but the consequence of the past, and the difference between natural and man-made things is merely between those whose potentialities necessarily grow into actualities and those that may or may not be actualized. Under these circumstances, any notion of the Will as an organ for the future...was entirely superfluous.\(^{538}\)

Arendt suggests in fact that this explains *why* any notion of the will more substantive than mere voluntary action (intentional as opposed to unintentional acts) is missing from Greek philosophy.\(^{539}\) This "curious lacuna...is in perfect accord with the time concept of antiquity, which identified temporality with the circular movements of the heavenly bodies and with

\(^{536}\)Ibid, Part Two, pp.13-14.


\(^{538}\)Ibid, p.15.

\(^{539}\)The absence that, as we saw in Chapter I, Bernard Williams attributes to the creation of artificial expectations in modern moral philosophy.
By contrast, the notion of will as a capacity "that in principle is indeterminable and therefore a possible harbinger of novelty" fits with the notion that the entire world had a "divine beginning," and is therefore a radical novelty. It is against the background of the "rectilinear time concept" required by the story of Christ’s ressurection and the Christian eschatology that, Arendt says, "the Will and its necessary Freedom in all their complexity were first discovered by Paul" [the apostle]. And she adds that the continuing resistance to acknowledging this phenomenon, manifested in philosophy from the medieval period to the present, shows how strong the reluctance was "to grant human beings, unprotected by any divine Providence or guidance, absolute power over their own destiny and thus burden them with a formidable responsibility for things whose very existence would depend exclusively on themselves..." Here we see Arendt at her most existentialist: she reads every reduction of the will, including the German idealist tendency to "equate Willing and Being," and the refusal of free will in Nietzsche and the later Heidegger, as episodes of bad faith.

Whatever we might make of her historical analyses of this issue, the innovation in Arendt’s own analysis of willing lies in her development of the theme of novelty. For Arendt, the motivational power of volition, its surplus over desire, is found in its conferring practical meaning on an indeterminate future. In contrast with liberum arbitrium, which

---

540Ibid, p.16.
541Ibid, p.18.
544And there would be much to say about this: see her discussions of Hobbes, Spinoza, Schopenhauer, Ryle, Wittgenstein, Kant, and Descartes, which I will not try to summarize here: Ibid, pp.23-28. I find Arendt very convincing, especially regarding the motives for the continual evasion of the will, although I recognize that many others will not.
simply picks from pre-ordained options, "spontaneity" as Arendt understands it is "a power to begin something really new" which "could not very well be preceded by any potentiality."\textsuperscript{545} She allows that the end the will gives itself is \textit{logically} possible, but denies that it is the agent’s \textit{potential} in the sense of something that is teleologically anticipated and simply unfolds. Although she does not contrast the will’s \textit{creation} of new ends with the relative passivity of desires, Arendt forces us to see how implausible it is that all the ends we take up could derive simply from the prior telos of our well-being: "Can anybody seriously maintain that the symphony produced by a composer...existed in a state of potentiality, waiting for some musician who would take the trouble to make it actual?"\textsuperscript{546} The sheer complexity and variety of \textit{products} of human willing, "art objects as well as use objects,"\textsuperscript{547} attests to the fact that through will, we conceive goals for ourselves for which there is not necessarily any \textit{prior impetus}; we conceive new challenges and invent new needs for ourselves. It is clear how well this fits my description of projective motivation: through projection, we can \textit{work up} our own motivation, sometimes over a period of time. Only understanding the will in these terms allows for the innovations through which historical time becomes \textit{linear}:

It is not the future as such but the future as the Will’s \textit{project} that \textit{negates} the given. In Hegel and Marx, the power of negation, whose motor drives History forward, is derived from the Will’s ability to actualize a project; the project negates the now as well as the past and thus threatens the thinking ego’s enduring present.\textsuperscript{548}

I think Arendt is right that nothing truly novel in human affairs could come about, except by sheer accident, if projective motivation were impossible. For the instinctual

\begin{flushright}
\textsuperscript{545}Ibid, p.29. \\
\textsuperscript{546}Ibid, p.30. \\
\textsuperscript{547}Ibid, p.31. \\
\textsuperscript{548}Ibid, p.36.
\end{flushright}
appetites vary little; the fundamental conditions of overall human well-being remain constant, altering only in the means necessary for their realization in different historical contexts; and brute preferences are too subjective or inarticulate about their reasons to extend to complex and incommensurable options: nobody just finds themselves ‘prefering’ to compose a symphony in a new style rather than to watch sitcoms, in the way that they prefer chocolate. We don’t ‘have a hankering’ to do a symphony; except in cases that would border on pathological, we can have such a goal only by making it our project. Of course we can project ends that are absolutely mundane and repetitive, but precisely because the will’s resources are greater than this, if we do so, the suspicion naturally arises that what we are really projecting is the goal of living in a way that requires such modesty in our pursuits (which itself could be based either on good or bad grounds).

Arendt does not concentrate on the distinction between act and motive, but her remarks on the "tonality" of the will reveal her understanding of the distinct kind of motivation it involves. There is in Arendt’s view a fundamental tension between the mind’s two aspects: thinking and willing. She quotes Bergson’s suggestion that "In the perspective of memory, that is, looked at retrospectively, a freely performed act loses its air of contingency under the impact of now being an accomplished fact..." In thought, which is hindsight, our will’s projects seem to flow out of a preceding series continuously, as if they were necessitated. Thus the temptation to reduce the will to something else, although from

---

549 Someone trying to rebuild a shattered life may project for herself goals such as getting up the next morning on time, getting to work on time, eventually being able to buy a home, "settle down" (whatever that means exactly) and live a halfway normal life — and she may have very good grounds for such projects. For her, this is courage. Similarly, someone joining a monastery might devote himself to attaining a series of little steps designed to build a lifestyle of peace. Someone else might be trying to fit in better with the alien culture in which they now live. But there are bad reasons as well: someone might be trying to prove to themselves that they are too inferior or worthless to attempt anything great.

550 One might compare these two aspects in her account to the objective and subjective points of view in Thomas Nagel’s account (see my discussion of Nagel in the forthcoming chapter on liberty of identification).

"the perspective of the willing ego, it is not freedom but necessity that appears as a delusion of consciousness." This "clash between thinking and willing" is also a matter of the opposing tones or moods of these mental activities. Thinking, even in working out difficult questions, is self-contented:

Rememberance may affect the soul with longing for the past, but this nostalgia, while it may hold grief and sorrow, does not upset the mind’s equanimity, because it concerns things which are beyond our power to change. On the contrary, the willing ego, looking forward and not backward, deals with things which are in our power but whose accomplishment is by no means certain. The resulting tension, unlike the rather stimulating excitement that may accompany problem-solving activities, causes a kind of disquiet in the soul easily bordering on turmoil, a mixture of fear and hope that becomes unbearable when it is discovered that, in Augustine’s formula, to will and to be able to perform, *velle* and *posse*, are not the same....

Speaking in terms of tonality—that is, in terms of the way the mind affects the soul and produces its moods, regardless of outside events, thus creating a kind of life of the mind—the predominant mood of the thinking ego is serenity, the mere enjoyment of an activity.... The predominant mood of the Will is tenseness, which brings to ruin the "mind’s" tranquility.

As argued in Chapter II, we could describe the "tone" of desire as a state of unquiet or dissatisfaction attempting to return to harmony and inner equilibrium. What differentiates desire from will in this respect is that in volition, the ego willingly takes on the tension, disturbance, and risk of having an unrealized end. By willing, we voluntarily move ourselves from a state of relative pose and quietude into a state of interestedness that risks frustration and disappointment.

Arendt means to emphasize this point when she writes

...in flagrant contrast to thinking, no willing is ever done for its own sake or find its fulfillment in the act itself. Every volition not only concerns [contingent] particulars but—and this is of great importance—looks forward to its own end, when willing something will have changed into doing it. In other words, the

552Ibid, p.31. She approving cites Duns Scotus as anticipating Bergson on this point.
553Ibid, p.34.
normal mood of the willing ego is impatience, disquiet, and worry (\textit{Sorge})...because the will’s project presupposes an I-can that is by no means guaranteed.\footnote{Ibid, p.37. Note Arendt’s allusion to Heidegger’s notion of “care” (\textit{Sorge}) as the being of Dasein.}

It is important to qualify this statement, since it could be misread. Arendt is not here asserting that the ground of willing must lie in the perceived value of the end to be achieved (and we have seen that a volitional state may take its own value as its ground). Rather, Arendt is pointing out a basic difference between all motivational states—including projective ones—and other non-motivational psychic states, such as pure cognition, which can be conducted in a mood that is practically speaking neutral. All motivation is focused on an end that must still be realized in the world, and thus seeks to fit the state of the world to its intentional content. Projective motivation is distinguished by the fact that it freely undertakes the worry and tribulation that motivation involves; it \textit{makes itself} dissatisfied with the actual states of affairs, or brings about an internal "restlessness"\footnote{Ibid, p.44.} quite similar to that from which desire always \textit{starts}.

Why then would anyone do something so crazy? There have been different answers. Nietzsche proposes a fundamental drive to ascendance, to distinction, to greatness in comparison with others. But human experience shows that ends having little relation to this result can be willed as well. In this chapter and the previous one, I have suggested several different possible grounds for projective motivation, and it is perhaps time to draw up a preliminary summary of them, however incomplete we may later find it to be.

\textbf{Some of the Possible Grounds of Projection}

1. The Moral Value of an end E irrespective of the positive or negative impact of realizing this end on agent’s flourishing.
   1a. End E is required under a formal or deontic standard of universalizability of impartial fairness or justice;
   1b. End E is inherently good according to non-formal ethical standards; ---e.g. E is the good of another person whose inherent value is recognized in
love.

2. The Aesthetic Value of an end E irrespective of its value for the agent’s well-being;
   2a. E is an end whose realization would be beautiful, or inherently contribute to the
       harmony of the world, or communicate something important to others;
   2b. E is an end whose realization would advance the ends of a tradition of
       practice in which the agent participates;
   2c. E is an end whose realization is difficult or challenging and thus presents an
       inherent opportunity for projective motivation.

3. The Ethical value of pursuing an end E irrespective of the positive or negative impact of
   that activity on the agent’s flourishing;
   3a. Taking an interest in E or caring about E is morally required of persons in the
       situation in which the agent finds himself or herself;
   3b. Devoting oneself to E, whether E is realized or not, tends to realizes goods
       internal to practices to that connect us to living traditions of human activity
       and foster a sense of communal identity (which provides a reason to join the
       practice);
       —e.g. projecting an end E on the basis of one’s heritage.
   3c. Devotion to E is itself an inherently worthy state according to nonformal criteria
       of moral worth, such as the duty of agape;

4. The Personal value of pursuing an end E, irrespective of the impact of commitment to it
   on the agent’s well-being;
   4a. Commitment to E provides a sense of purposefulness and is part of a pattern of
       caring that generates existential meaningfulness in the agent’s life;
   4b. Commitment to E is demanded by an inscrutable sense of higher ‘calling;’
   4c. Commitment to E is required by, or at least fits with, other prior commitments we
       have made and the shape of our live to date (‘backward looking’ reasons); —e.g.
       E is valued and pursued by friends to whom we have committed ourselves;
   4d. Commitment to E developed during our pursuit of E for other reasons, such
       as its power to satisfy various desires, and now this commitment is an
       established part of our character.

This division of grounds for projection into aesthetic, ethical, moral, and personal
reasons is not meant to be exhaustive: aside from tradition and culture, mere social
convention and popular opinion can be taken as grounds for projective motivation; aside
from the personal grounds mentioned, others such as religious background, revelations,
traumatic events, and many other distinctive experiences not catalogued here can become the
bases for the resolve in which a person steels himself projectively to new ends. A fuller
account than I can give here would explain how in the temporal process of human lives,
prior commitments (whether recognized or not) necessarily condition and color how various possible grounds for self-projection appear to us, and which we can even consider. As the medievals emphasized, there is a reciprocal relation between the development of volitional and cognitive aspects of one’s personality. This helps explain the possibility of all sorts of perverse, self-deceptive, or corrupt grounds for projective motivation: one can take the vulnerability of others as an opportunity to dominate them; one can project frustrating or even self-destructive ends for on the ground of deep self-loathing (as we saw with Sartre’s example in Chapter II).

As we will see in the next chapter, we must also consider grounds deriving from the fact that through forming higher-order projects concerning our own character and motives, we can shape the kind of person we become over time. Thus the possible grounds of projection also include the inherent value of the kind of self-control that is realized through what Harry Frankfurt calls "identifying" with some possible motives for outward action over others.
CHAPTER IV

A New Phenomenology of Frankfurtian Volitional Identification
Contents of Chapter IV

IV.1. Introduction: The Importance of Identification

IV.1.1. Foucault and Identification as ‘Care of the Self’

IV.2.1. The Existential Significance and Function of a Theory of Identification

IV.2. Frankfurt’s Initial Account of Volitional Identification in Terms of ‘Desires’

IV.2.1 Identification and Personhood

IV.2.2. The Initial Hierarchical Model

IV.2.3. Identification, Freedom, and Rationality

IV.3. Objections to Frankfurt’s Initial Account

IV.3.1. Two Initial Objections Against Identification as a Central Phenomenon

A1: The Animal Self Objection

A2: The Davidsonian Objection

IV.3.2. Five Objections Against the Bare Hierarchical Account of Identification

B1: The Origin of Identification Objection (i & ii)

B2: The Regress Objection

B3: The Ab Initio Objection

B4: The Sporadicity Objection

B5: The Arbitrariness Objection to Decisive Identification as ‘Resounding’

IV.4. Developments in Frankfurt’s Later Accounts of Identification

IV.4.1. Coercion and the Qualitative Argument Against Obj.B1

258
IV.4.2. Identification as Strongly Evaluative Decision

IV.4.3. Care and Decision

IV.5. A First Approach to Decisive or ‘Wholehearted’ Identification

IV.5.1. The Proper Function of an Account of Decisive Identification

IV.5.2. Decisive Identification as Volitional Necessity or Satisfaction
   -- Volitional Necessity
   -- Volitional Satisfaction
   -- Three Problems with the Satisfaction Account

IV.6. The Projective Account of Identification

IV.6.1. Three Relevant Orders of Identification and the ‘Personal World’
   (A): Identification₁: As Modally Extended
   (B): Identification₂: Penelhum and Frankfurt
   (C): Identification₃: The Highest-Order Will

IV.6.2. Identification as a Distinctively Volitional Type of Motivational State

IV.6.3. Conclusion: Heidegger and Authentic ‘Care of the Self’
IV.1. **Introduction: The Importance of Identification**

In the Introduction to this dissertation, we saw that in contemporary philosophy of mind, it is generally assumed that to be a *person* requires (at most) having certain types of consciousness, rationality, and capacity for intentional action.\(^{557}\) As we have seen in Chapters I and II, however, an existential analysis of human agency undermines this assumption by emphasizing both our capacity for *decision* (second-order acts in which we form first-order intentions) and our capacity for *projective motivation* in making such decisions, i.e. to guide our intention-formation by ends that we *give ourselves*, as opposed to passively desiring or being attracted to them for their potential to satisfy our preferences or fill gaps in our well-being. We have also seen, in Chapter III, that there is a long philosophical tradition, stretching at least from Scotus through Kant to contemporary writers such as Levinas and Frankfurt, whose writings on ethics and moral psychology explicitly or implicitly recognize this capacity for projective self-motivation.

In this chapter, this preceding analysis of projective motivation will be linked with phenomena which Frankfurt has called "identification." Of course, this term has many senses. We speak of "identifying" someone *as* the person who performed some feat, or who is related to someone else in a given way. Identification in this sense refers to picking out some social or historical role an agent plays. We speak of "identifying" *with* someone in the sense of feeling empathy for them, feeling that we understand what it is like to be in their position, and so on. We also speak of identifying with a role model or character in a novel:

---

\(^{557}\)Whenever I use ‘action’ by itself, I mean ‘action’ in the first-order sense, i.e. acts like remembering a face or walking to the refrigerator, which are directed by intentions aimed at ends *other* than forming intentions.
we then refer to feeling attracted to (or approving of) what that character represents. But "identification" in the specifically volitional sense is a term of art referring to an intrasubjective relation in which the person adopts, appropriates, or accepts certain of her desires, emotions, and reasons for acting as motives or ends that express who she wills to be. "Identification" in this special sense refers to these motives or ends of action being her own in a way that goes beyond, for example, her consciousness of them as actual or possible motives, or her acting on them with reflective awareness that they inform her intentions. When she identifies with them, these motives are related to her agency in a further more intimate or personal sense. That there is such an ‘inward’ relation of personal incorporation of motives is the genuine mystery to which the term "volitional identification" refers, and which analyses of identification in this philosophical sense must try to explain.

In this chapter, I will argue that the development of this concept in Harry Frankfurt’s work points towards the need to interpret identification as a kind of self-regarding state of projective motivation. The notion of projective motivation helps resolve longstanding problems in the analysis of identification, while, in turn, the phenomena of identification evince the reality and significance of the hypothesized projective structure. The next chapter will confirm this by showing that, while the main alternative approaches resolve recognized problems in Frankfurt’s early account of identification, they all have other problems avoided by the existential analysis in terms of projection sketched in this chapter.

558 There may, however, be some relation between this specifically psychological sense of "identification" (as the term is used in personality development theory) and the phenomena of volitional identification that concern philosophers. Howard Kamler, for example, seeks to bridge this gap in his book, Identification and Character: A Book on Psychological Development (SUNY Press, 1994).

559 That is, reasons which figure in first-order intentions to perform actions.

560 And henceforth, when I use this word by itself, I always mean volitional identification in the special philosophical sense sketched here.
IV.1.1. Foucault and Identification as ‘Care of the Self’

Like the idea of second-order agency, the general notion of an "identificatory" self-relation has a long heritage tracing back to the Stoics and Augustine. Something like this kind of intrasubjective relation to one’s motives is implicit in the Stoic idea of the "care of the self." As Foucault argues, what distinguished Stoic concerns about sexual appetites was not a greater emphasis on "conjugal fidelity" than their Greek forbearers:

Rather, what stands out in the texts of the first centuries [AD] —more than new interdictions concerning sexual acts— is the insistence on the attention that should be brought to bear on oneself; it is the modality, scope, constancy, and exactitude of the required vigilance;...it is the importance attributed to self-respect [and] an intensification of the relation to oneself by which one constituted oneself as the subject of one’s acts. ²⁶¹

Foucault notes that this theme has a long history in the literature on virtue, being initially "consecrated by Socrates" who presents himself "as a master of the care of the self" in Plato’s Apology.²⁶² Similarly, "Taking care of one’s soul was a precept that Zeno [the Stoic] had given his disciples from the beginning," and it was developed in Seneca, who "commands a whole vocabulary for designating the different forms that ought to be taken by the care of the self..."²⁶³ But it is especially notable, in view of his contribution to the idea of projective motivation (see Chap. I.3) that Epictetus provides "the highest philosophical development of this theme," according to Foucault:

Man is defined in the Discourses as the being who was destined to care for himself. This is where the basic difference between him and the other creatures resides. The animals find "ready prepared" that which they need in order to live ... Man, on the other hand, must attend to himself: not, however, as a consequence of some defect that would put him in a situation of need and make him in this respect inferior to

²⁶²Ibid, p.44. For Socrates has thought critically about his own ignorance, and constantly examines whether his actions make his soul more or less ‘earthly,’ as well as listening to challenges from his daimon.
²⁶³Ibid, p.46.
animals, but because the god [Zeus] deemed it right that he be able to make free use of himself; and it was for this that he was endowed with reason...[which] enables one to use, at the right time and in the right way, the other faculties....The care of the self, for Epictetus, is a privilege-duty, a gift-obligation that ensures our freedom while forcing us to take ourselves as the object of all our diligence.\textsuperscript{564}

It is remarkable how clearly this Stoic theme prefigures Frankfurt’s groundbreaking thesis, advanced in his 1971 essay, "The Freedom of the Will and the Concept of a Person," that it is the capacity for reflexive or higher-order volition, through which we identify with ‘desires’ or motives for acting, which distinguishes human persons from animals (which are capable at most of first-order agency):\textsuperscript{565} And in this light, we should see Frankfurt’s treatment of volitional identification and related topics (such as personal ‘care’— Chap. III §2) precisely as an attempt to give a more detailed explanation of the same phenomena which Foucault interprets under the heading, ‘care of the self.’ As Frankfurt says in a recent essay, in addition to the ideal of freedom (in the sense of "varied alternatives available in the design and conduct of his life")\textsuperscript{566}:

Another ideal also enjoys considerable vitality... This is the ideal of individuality, construed in terms of the development of a distinctive and robust sense of personal identity. To the extent that people find this ideal compelling, they endeavor to cultivate their own personal character and styles and to decide autonomously how to live and what to do.\textsuperscript{567}

"Individuality" or "autonomy" in this sense is similar to Foucaultian "care of the self."

Frankfurt uses "autonomy," as distinguished from "freedom" (i.e. liberty\textsuperscript{568}), for an agent’s

\textsuperscript{564}Ibid, p.47. In this regard, also see Heidegger’s discussion of the ‘Myth of Care’ in \textit{Being and Time}, \S 42, H197-199, p.241-243, and my discussion of Heidegger at the end of this chapter.

\textsuperscript{565}Harry Frankfurt, "Freedom of the Will and the Concept of a Person," \textit{Journal of Philosophy}, 68.1 (January, 1971); reprinted in \textit{The Importance of What We Care About}: 11-25.


\textsuperscript{567}Ibid, p.17.

\textsuperscript{568}Since the term "freedom" is so ambiguous, and is used by other authors to refer to what Frankfurt here (continued...)
achievement of forming her mind and will not simply by imitation of others, but by a reflexive evaluation of motives that are already in a weak sense ‘her own.’

Foucault links this reflexive or inward focus on one’s own character with the new "individualism" of the Hellenistic and Roman world. But he notes that "individualism" is a term with three different connotations:

(1) ...the absolute value attributed to the individual in his singularity; "(2) "positive valuation of private life, that is, the importance granted to family relationships... [and] (3) the intensity of the relations to self, that is, of the forms in which one is called upon to take oneself as an object of knowledge and a field of action, so as to transform, correct, and purify oneself, and find salvation.

Foucault argues convincingly that these three aspects are separable: whereas 19th century bourgeois culture emphasized the private life, its individualism was weak and its sense of self-relation undeveloped; by contrast, the Stoic notion of personal virtue did not valorize private life but emphasized the third aspect: "cultivation of the self." As he says, "there are societies or groups in which the relation to self is intensified and developed without this resulting...in strengthening of the values of individualism or private life." And thus among the Stoics, the self-cultivation became "not an exercise in solitude, but a true social practice," which was partly institutionalized and depended on communicative exchange with

---

568(...continued)
calls autonomy, I will use "liberty" to refer to freedom in the sense of being able to bring about or having unmediated access to relevantly defined alternate possibilities. 'Liberty' is thus distinct from merely hypothetical freedom of alternatives, which conditions the agent’s "doing otherwise" on his intending or willing otherwise, without specifying that the agent is able to bring about these conditions.

569Ibid, p.17 and p.23.

570Foucault, Care of the Self, p.42.

571In view of Charles Taylor’s analysis of Rousseau and other thinkers in Sources of the Self, however, we might be forced to regard Foucault’s diagnosis of this culture as too simplistic.

572Foucault, p.43.

573Ibid.
others,\textsuperscript{574} which resulted in an "intensification of social relations" in spiritual friendships.\textsuperscript{575}

To make sense of Foucault’s distinction, we need to understand such self-cultivation in terms of the kind of reflexive volitional relation that Frankfurt calls "identification." As we will see in this chapter, any account of identification must specify three things: the type of attitude the identifying agent takes towards some inner ‘object’ or aspect of himself; the intensional object of the agent’s attitude; and the process by which this attitude comes about or is enacted. On my account, these three different variables will be interpreted roughly as follows.

(1) \textit{Attitude}. Identification is a relation in which the person determines what types of desires, motives, and dispositions he should be moved by in his ordinary (first-order) actions: to the extent that his actual motives differ from this determination, he is critical of them. As I will argue in this chapter, however, although the agent’s determination may be informed by ethical considerations (as for the Stoics it was influenced by the ideals of independence and limitations on the pursuit of goods subject to chance or fortune) it cannot simply amount to judging that a given type of motive \textit{ought to} move one, nor does it reduce simply to \textit{wanting} to have a certain motivational character in any desiderative sense. Rather, a person identifies with certain motives \((M)\) —and thereby also usually alienates opposed motives \((\neg M)\) in conflict with \(M\)\textsuperscript{576}— only by giving herself the \textit{project} of developing these motives as stable features of the ‘outward character’ she expresses or exhibits in ordinary

\textsuperscript{571}Ibid, p.51. On reflection, Aristotle’s attitude that his \textit{Nicomachean Ethics} is not for the young but rather for people who already practice virtue, in order to give reflective depth to their practices, already suggests the beginning of this genre of discussions of the soul. Yet Aristotle lacks the qualified optimism about human self-reformative capacities which characterize Stoic and later conceptions of virtue and care of the self.

\textsuperscript{575}Ibid, p.53.

\textsuperscript{576}I say ‘usually’ here because, as we will see, the possibility of tension within the higher-order will itself is crucial to understanding the nature of volitional identification.
actions. The relevant type of attitude is itself volitional in the distinctive middle-soul sense explained in Chapter II.

(2) Object. As just suggested, the intensional object with which we identify is not a fine-grained temporally specific motive, such as the feeling of exhaustion that motives an immediate desire to sleep. Rather, it is an entire pattern of such motives sustained over time. Such an ‘outward character’ consists in the dispositions and desires which normally motivate one’s intentions in observable courses of action: e.g. saying something nasty out of jealousy, or re-running a medical test out of diligence, or planning a party out of a desire to see old friends. ‘Outward character’ includes the kind of appetitive, emotional (or affective), prudential, and conscientious motivational states which others routinely attribute to us (more or less accurately) in interpreting our first-order intentions and ends in order to make sense of our actions. Such interpretation is constantly ongoing, and without it, the ordinary intercourse of interhuman life could not continue.577 By identifying motives $M$ that constitute an ‘outward character’ in this sense (as opposed to identifying with a specific short-term desire, like craving an ice cream cone while at the mall), the agent acquires an ‘inner’ or higher-order character, which consists in her enduring commitments to forming her intentions primarily in accordance with ends guided by $M$-motives. Since the motivational character with which someone ‘identifies’ in this fashion certainly need not be egoistic nor oriented primarily towards the ‘private’ sphere of family life, the general concern for volitional self-cultivation is not necessarily privatistic, nor does the agent’s volitional relation to ‘himself’ necessarily make him psychologically introverted, self-

---

577MacIntyre gives a good description of this aspect of agency in After Virtue, Second Edition (University of Notre Dame Press, 1984), Chapter 15, where he argues that actions differ from mere behaviors in that others are able to make them intelligible as actions by recognizing them as enactments of intentions the agent has (see pp.208-211). However, it is important to see that MacIntyre’s analysis of human acts here does not extend beyond first-order agency.
absorbed, narcissistic or especially self-reliant or "individualistic" in his social values.

These states have to be distinguished from caring for self.

This point is all the more significant because many authors have diagnosed supposedly prevalent problems —e.g. lack of meaning and virtue in the twentieth century— as the result of an egoistic individualism lacking the idea that one must "take care of oneself" in the Stoic sense of forming one’s character. This concern provides an ethical subtext for Frankfurt’s analysis of the "wanton" as a human being who omits volitional identification and thus fails to take a stand on what kind of person he wants to be. The theme of self-cultivation suggests that identification is a proactive form of ‘self-care’ or concerned involvement with the formation of our own motivational structures. Thus the concept of identification is inevitably related to the classical themes of virtue and responsibility for one’s moral character.

(3) Process. Finally, on my account, the distinctively volitional attitude towards one’s outward character that constitutes identification must itself be acquired through a process that involved constrained but real libertarian freedom. We will see that on this issue, my account varies from Frankfurt’s and most others. On the type of attitude question (1), my account agrees with Frankfurt’s care theory, but disagrees with the other types of attitude he has proposed at various stages in his authorship, and with the more rationalist proposals of several other theorists. On the question of object (2), my approach agrees with Frankfurt’s

---

578 Ibid, p.43. For example, see Charles Taylor’s diagnosis of the “malaises” of modernity in *Ethics and Authenticity* (Harvard University Press, 1991). These malaises include "individualism," which has led to egoistic selfishness (p.4), "the primacy of instrumental reason" (p.5), and the resulting decay of social structures and loss of the civic virtue necessary for democratic institutions to strengthen both fraternity and citizens’s commitment to the common good (pp.8-10).

579 This was for a long time part of the idea of a liberal education. As Foucault notes, "Educating oneself and taking care of oneself are interconnected activities. Epictetus lays stress on this point: he does not want his school to be considered as just a place...where one can acquire knowledge useful for a career" (*Care of the Self*, p.55).
IV.1.2. The Existential Significance and Function of a Theory of Identification

In analytic moral psychology today, however, Frankfurt’s work on identification and the "higher-order will" is not usually connected with the Stoic themes of self-care and virtue, but is interpreted almost solely as an attempt to defend a compatibilist account of the autonomy or self-determined status of first-order acts. Following his famous 1969 paper which attempted to show that there are counterexamples to the "Principle of Alternate Possibilities" for moral responsibility, it was assumed that Frankfurt developed his "hierarchical" conception of identification in subsequent papers solely in order to show how compatibilists could handle cases where persons act on desires produced by addiction, threat, hypnosis, or other types of manipulation.

Yet while Frankfurt’s work has profoundly affected our understanding of when human acts are "free" in the sense necessary to make us morally responsible for them, and while he has undoubtably approached this issue from a generally compatibilist perspective, he...

---


582Such cases created obvious difficulties for what Shatz calls "classical" compatibilist accounts, like Schlick’s or Moore’s, which generally held that a person S does action A freely if "in doing A, S is acting in accordance with and because of his motivation—he is doing as he wants, prefers, chooses, wills, decides, tries, or intends with respect to A..." (Shatz, p.451-2). But Frankfurt introduces his hierarchical distinction to show how a person may act freely on compulsive desires if he identifies with them, or not if he fails to identify with them, thus extending the compatibilist idea that the free/unfree distinction can still be drawn solely in terms of the agent’s motivational states (Shatz, p.460).
suggests that it is not merely "for the sake of moral responsibility that we care as much as we do about being free," and maintains that

my philosophical attention has for the most part been guided less by an interest in questions about morality than by a concern with issues belonging more properly to metaphysics or to the philosophy of mind — for instance, how we are to conceptualize ourselves as persons, and what defines the identities we achieve.

In this respect, Frankfurt’s primary focus is on the same questions that have dominated two closely related traditions of continental thought known as "existentialism" and "personalism" respectively. On my reading, the main significance of Frankfurt’s work on volitional identification lies in this direction, in its ability to clarify the essentially intrapersonal or individual character of persons as self-shaping selves.

To appreciate the importance of Frankfurt’s theme of identification for existential thought, note that all the following are central in Kierkegaard’s existentialism, and to a large extent also in the writings of Jaspers and Heidegger, though only to a lesser extent in Sartre:

1. The idea of depth or inwardness: the ‘real self’ underlies the outward roles a person plays in social life, hidden behind the surface level of desires and inclinations which figure in the person’s everyday acts and his public persona;

[583]Frankfurt, The Importance of What We Care About, Preface, p.viii.


[585]"Personalism," beginning with Max Scheler’s Formalism in Ethics and Non-Formal Ethics of Values, and including works such as Emmanuel Mounier’s Personalism, Reinhold Niebuhr’s The Nature and Destiny of Man, William Sadler’s Existence and Love, Jacques Maritain’s The Person and the Common Good, and arguably parts of Buber’s and Levinas’s work, has generally been more theistic and more focused on the theme of agape than philosophical existentialism in general — although both have a common root in Kierkegaard. Here again, Frankfurt has some affinity to these themes, as his more recent focus on types of love suggests.


[587]Note that the following list is not meant to count as a definition of ‘existentialism,’ which is in any case a diverse family-resemblance concept at best, but to single out several ideas in the Kierkegaardian tradition which converge to an interesting degree with Frankfurt’s conception of the self.

[588]Consider, for example, Heidegger’s famous critique of the public or "they"-self in Being and Time, ¶27.
(2) The idea of self-integration through ‘choosing oneself,’ taking responsibility for one’s character and what one makes of the situation and tendencies one has by birth and social environment: the person “purifies” his character by guiding his surface level motives and interactions with others according to the highest
consider for example Kierkegaard’s Purity of Heart is to Will One Thing, tr. Douglas V. Steere (Harper Torchbooks, 1958).


Consider, for example, Kierkegaard’s discussion in “Subjective Truth, Inwardness; Truth is Subjectivity” in his Concluding Unscientific Postscript, tr. Howard V. Hong and Edna H. Hong (Princeton University Press, 1992): 189-250.

Though these theses are not in any straightforward way identical to Frankfurt’s, it must be acknowledged that in his focus on volitional identification, Frankfurt treats many of these issues that have occupied existentialist writers.

However, if existentialists can look towards an analysis of volitional identification for insight into the structure of individual identity, it will be in part because (a) it is essential to persons that they normally be capable of moral responsibility not only for their immediate bodily (and mental) acts, but also for the abiding character they exhibit in the course of such actions, which is so closely connected to their relations with others; and (b) volitional identification (and the form of the self which it disclose) is involved in the explanation of these moral capacities. As John Fischer notes, for Frankfurt a "person" is "accessible to the reactive attitudes," or subject to judgments of moral worth, only because he is capable of commitments of his inward self.589

(3) The idea of a ‘center’ from which the choices and acts of self-integration radiate: although personal existence is finite and always enacted in relation to a world of possibilities limited by facticities, historical conditions, and relations with others, it also transcends these conditions insofar as its own free projects and purposes determine the meaning of its identity within these relations.590

(4) The idea that ‘will’ is more distinctive of personhood than objective reason: persons’ subjective commitments and choices give their existence an irreducibly unique significance which shared forms of reasoning cannot determine, and no objective criteria can capture.591

---

589Consider for example Kierkegaard’s Purity of Heart is to Will One Thing, tr. Douglas V. Steere (Harper Torchbooks, 1958).


591Consider, for example, Kierkegaard’s discussion in "Subjective Truth, Inwardness; Truth is Subjectivity" in his Concluding Unscientific Postscript, tr. Howard V. Hong and Edna H. Hong (Princeton University Press, 1992): 189-250.
Like others, Fischer has challenged this claim and has developed an alternative actual-sequence account of moral responsibility for particular actions (though not for character) which dispenses with the notion of identification. But this is because Fischer finds identification-based accounts, or "mesh theories" as he calls them, unable to yield convincing conditions for moral responsibility. This accurately presupposes, however, that it is a condition on any successful account of identification that it be relevant to the conditions for moral responsibility and fit coherently into an account of such responsibility. Thus no explanation of identification can just sidestep Fischer’s challenge.

Nevertheless, the problem with Fischer’s approach, like most of the current literature on autonomy, is that it proceeds as if facilitating an account of moral responsibility were the only relevant condition for an adequate account of volitional identification. Partly to correct this imbalance, in this chapter, I will not approach the analysis of identification just as a step in a theory of moral responsibility, as Fischer has, but consider it phenomenologically, or ‘in its own right.’ So considered, we will find that identification can shed light on existential themes concerning the nature of individual selfhood, but only once we both (a) revise Frankfurt’s analysis in ways that distance our understanding of identification from its early association with Humean moral psychology, and (b) ultimately separate the notion of identification from its compatibilist roots (for there is nothing essentially compatibilist about

595 I hope to place the notion of identification in the broader context of an account of moral responsibility and consider its relation to responsibility for one’s character in the projected additional chapter on freedom of the higher-order will.
Frankfurt’s phenomenology of identification).\textsuperscript{596} The first revision is completed in this chapter and Chapter V, while the second is left for an additional chapter on freedom of the higher-order will, which had to be left out of the dissertation for space reasons.

**IV.2. Frankfurt’s Initial Account of Volitional Identification in Terms of Desires**

**IV.2.1 Identification and Personhood**

The paper in which Frankfurt first explicitly thematized volitional identification\textsuperscript{597} had its immediate context in Sir Peter Strawson’s claim that our cognitive capacities and "corporeal characteristics" taken together could be regarded as sufficient for personhood, since they ground an irreducibly ‘individual’ form of self-reference.\textsuperscript{598} Despite its Kantian source, Strawson’s theory reflected the still-dominant ‘mentalist’ presupposition that the mind/body relation is the main problem in understanding personhood, "rather than the quite different problem of understanding what it is to be a creature that not only has a mind and a body but is also a person."\textsuperscript{599} This statement signals Frankfurt’s decisive break with what I have called the mentalist tradition going back to Plato and Aristotle (see Int. §I)—a step which takes him, as I will argue, in a direction that only a fully existentialist account of personhood can fulfill. In "Freedom of the Will and the Concept of a Person" and later articles which develop its themes, Frankfurt argues that the feature which distinguishes persons from other

\footnotesize
\begin{itemize}
\item \textsuperscript{596}In "Conditions for Freedom of the Higher-Order Will: Frankfurt and Augustine" (presented at the Eastern APA, Atlanta: December 1996), I argued that Frankfurt’s account of free will and identification is at least compatible with a libertarian or indeterministic understanding of alternate possibilities of identification itself, such as the one implicit in the early writings of Augustine.
\item \textsuperscript{597}Harry Frankfurt, "Freedom of the Will and the Concept of a Person," Journal of Philosophy, Vol. LXVIII, No.1 (January 14, 1971); reprinted in The Importance of What We Care About: 11-25.
\item \textsuperscript{598}See P.F. Strawson, Individuals (Methuen, 1959). Strawson’s account is at least partially inspired by Kant’s focus in the Critique of Pure Reason on our mental capacities for judgment and self-awareness rather than on our volitional capacities.
\item \textsuperscript{599}Frankfurt, "Freedom of the Will and the Concept of a Person," p.11, note 1, my italics.
\end{itemize}
higher animals is not levels of consciousness, nor even advanced capacities for deliberation and reasoning, but rather "the structure of a person’s will" which gives them the capacity for volitional identification. In his recent essay "On the Necessity of Ideals," Frankfurt repeats this point, noting that an individual’s "personal characteristics" are "characteristics of his will" rather than of his "nature as a human being or as a biological organism of a certain type." This also puts him at odds with the classical notion that persons are individuals of the species ‘rational animal’:

Reason has usually been regarded as the most distinctive feature of human nature and the most sharply definitive. I believe, however, that volition pertains more closely than reason to our experience of ourselves and to the problems in our lives that concern us with the greatest urgency.

However, since the current literature analyzes Frankfurt’s model of the will mainly for what it may contribute to our understanding of autonomy, rather than the animal/person distinction, the uniquely personal capacity for volitional identification has been understood simply as a feature of what Williams calls "mature agents:" for autonomy represents an "ideal of maturity as responsibility for [one’s] self," which goes beyond the voluntariness thought sufficient for responsibility for particular acts on classical accounts. This may not have been Frankfurt’s initial intention, but today, identification is used in explaining the

---

600 Ibid, p.12.
601 Ibid, p.18.
603 Frankfurt, The Importance of What We Care About, Preface, p.viii.
604 Bernard Williams, "Voluntary acts and responsible agents," in Williams, Making Sense of Humanity: 22-34, p.28.
605 Ibid, p.29.
606 Although Frankfurt’s notion of “freedom of will” or full autonomy goes beyond what is necessary for “acting freely” or being responsible for one’s acts ("Freedom of the Will," p.24), “acting freely” is also more than acting intentionally or voluntarily in Aristotle’s sense, since it also involves the capacity for identification (as Frankfurt’s papers on coercion make clear).
first of what Gary Watson calls the two primary senses of freedom, which are: "self-determination (autonomy) and the availability of alternative possibilities." In this chapter, I will focus strictly on identification as part (the key part) of self-determination or "autonomy," leaving the relation between identification and liberty for a projected additional chapter on free will.

Sections §2.2 - §2.3 will explain Frankfurt’s first proposed analysis of identification in terms of a hierarchy of desires. Subsequent sections will then (1) show why this initial model is inadequate to its task of explaining identification; (2) explain several innovations Frankfurt introduced which go some way towards resolving these problems; (3) critique Frankfurt’s two recent accounts of decisive identification, and (4) propose a new account of identification as a special kind of ‘projective’ volition and distinguish three degrees or ‘levels’ of identification that are integral to personhood. (5) I conclude by linking this projective account of identification with Heidegger’s analysis of human ‘ek-sistence’ and its potential for authentic self-care.

Chapter V then supplements this argument with a critical review of a representative set of alternative analyses of identification which others have proposed to resolve the problems with Frankfurt’s initial model. The problems with these alternative accounts of the phenomena helps explain why the projective reinterpretation points towards an existential conception of persons at odds with the rationalist, interpersonal, and pragmatist models of personhood reigning today.

---


608 As discussed in the Introduction §2.
IV.2.2. The Initial Hierarchical Model

Frankfurt begins his "Freedom of the Will" paper by defining an agent’s first-order "will" as "the desire (or desires) by which he is motivated in some action he performs or...the desires by which he will or would be motivated if or when he acts," i.e. the desire(s) that is actually operative in her actions or which will move her to do what she does when she acts. At this point, Frankfurt meant ‘desire’ in a very generic sense here: he did not distinguish between "desires and motives" or between "preferences and purposes;" and he noted that "The desires in question may be conscious or unconscious, they need not be univocal, and [we] may be mistaken about them." Hence we might restate his definition as saying that one’s first-order "will" (or "will,") is the motive (out of all one’s conscious motives) which is incorporated into one’s intention to act, or which provides the end E that one intends to further in doing some (first-order) act A. As Pink points out, desires by themselves are not actions (since "I can want [end] E, without wanting E to be brought about through my agency") whereas intentions are understood as "ensur[ing] that one is motivated outright to act as intended." Thus if I intend to A in order that E, I must either be trying to act on this intention, or in a state such that (as long as this remains my intent) I will (try to) act on it when the required enabling circumstances are fulfilled. It is one’s ‘desire’ for the

---

611 Ibid, p.12.
613 This definition fits Frankfurt’s account as long as we assume that the intention is effective. Frankfurt notes that "even though someone may have a settled intention to do X, he may nonetheless do something else instead of X, because his desire to do X proves to be weaker or less effective than some conflicting desire" (p.14). At first glance, this appears curious, because it suggests that there are cases in which we act on a desire (our "will,") without any intention so to act. This becomes comprehensible in light of Frankfurt’s later paper on action—discussed below—which explains that although all actions are intentional movements, only some are preceded by more deliberate intentions to act in a given way (or to perform an intentional movement.
end E that figures in one’s operative intention to act in this sense that constitutes one’s will for Frankfurt.615

Frankfurt also says of someone who only has desires at this level, "His actions reflect the economy of his first-order desires," which suggests that in such a being, the intention on which he acts is simply a function of which existent desire is strongest at the moment in his psyche. Thus at the first-order level, as Pink says, Frankfurt gives "the same account of the will as did Hobbes."616 Yet it is important (as we will see) that nothing in Frankfurt’s subsequent analysis hangs on the idea that first-order agents do not actively form intentions to act, or that they do not decide to perform an action A on the motive supplied by the same first-order desire to A. And in his own account of action, Frankfurt explicitly rejects the "causal theory in which actions are construed as essentially movements whose causes are desires and beliefs," because there are counterexamples where desires and beliefs cause movements without them constituting actions.617 What distinguishes actions from movements is rather that they are guided by the agent as she is performing them,618 which makes them "intentional movements."619 Some ‘actions’ in this sense are also intentional actions, in which the agent "more or less deliberately or self-consciously" intends "that

615 One’s "will," in this sense is similar to the Kantian notion of an action-maxim, which includes a description of the act to be done and the end or purpose for which it is done (a purpose which may be derived from preferences, desires or impulses, but which is incorporated into an intention and thus actually becomes operative in action).

616 Pink, The Psychology of Freedom, p.34.

617 Harry Frankfurt, "The Problem of Action," reprinted in The Importance of What We Care About: 69-79, p.70. Frankfurt describes a case where a man at a party wants to spill a glass "to signal his confederates to begin a robbery," and believes his spilling will accomplishes this, but gets nervous because of this desire and belief and spills the glass anyway out of nervousness. Frankfurt thus denies that "spasms and twitches" are actions if they are caused by beliefs and desires (see note 1, p.70).


619 Ibid, p.73.
certain intentional movements of his body should occur.”620 In these cases, his guidance of his movements is itself guided by his intention so to guide his movements.621 This account of deliberate or "intentional action" seems very close to Pink’s notion of deciding to act in a given way, or intentionally forming the act-constituting intention that will guide one’s movement. For Frankfurt, in some cases, then, the translation from first-order desire to act-constituting intention may involve a decision, although he denies that a movement or piece of behavior counts as an action because it is caused by the decision.

Frankfurt then defined a second-order volition as a desire to act on a certain first-order desire D, or to have D become our will. By contrast, a mere second-order desire is a desire merely to desire ‘to X,’ which leaves it open that we might not want the ‘desire to X’ to become our operative "will.”622 Second-order volitions are thus simply one sort of second-order desire, i.e. second-order desires with a special kind of content623: they are not merely about experiencing desires, but about acting them. A second-order volition is a desire to have another desire be the effective motive if or when one acts.

It is important to see, however, that the distinction between second-order desires and second-order volition(s) is not constructed simply in parallel with the distinction between first-order desires and the agent’s first-order will. For a second-order volition is not distinguished from other second-order desires by its effectiveness either in controlling the agent’s actual set of first-order desires, or in controlling which of these desires actually

620Ibid, p.73.
621Ibid, p.74.
622Ibid, p.15. Frankfurt’s example is a doctor who wants to experience the desires his drug-addicted patients feel, without these desires becoming so strong that he acts on them. In my terms, this doctor does not actually want to have drug-addiction become part of the ‘outward character’ expressed in his acts.
623As Frankfurt says later, "Someone has a desire of the second-order either when he simply wants to have a certain desire or when he wants a certain desire to be his will. In situations of the latter kind, I shall call his second-order desires ‘second-order volitions’” (p.16).
motivates the agent to act. It only needs to be a desire with the right content to be a second-order volition. An agent can have such a desire if the desire he desires to be effective in moving him is not the desire on which he ends up acting. Thus unlike other animals, persons "are capable of wanting to be different, in their preferences and purposes, from what they are."624

By contrast, the first-order will is effective by definition. If a motive is not the desire or set of purposes we act on, it is not our will. Of course, it may not be the set of motives that we want to be effective, but it is our first-order will in Frankfurt’s sense nonetheless. As a result, the first-order will is by definition unique: an agent can only have one will in any given act (or by extension only one pattern of will in any series of acts), whereas Frankfurt’s definition leaves it open in principle that an agent could have multiple (and thus conflicting) second-order volitions. But Frankfurt does not seem to perceive this clearly in his "Freedom of the Will" paper: thus he suggests that "If there is an unresolved conflict among someone’s second-order desires, then he is in danger of having no second order volition" in the singular, as if there could only be one "preference concerning which of his first-order desires is to be his will."625 In fact, however, people can and often do have conflicting attitudes about this matter. As we will see, this asymmetry in the "desire"/"volition" distinction between the first and second orders has implications which Frankfurt did not initially anticipate.626

---

626Note that we might be able to resolve this asymmetry if we assumed that a person can at most have one settled second-order intention about which first-order desires should move him—or only one intention about how to form his first-order motives for acting. Then we could say that the second-order volition was the motive incorporated in this intention. Something like this is involved in Frankfurt’s later notion of decisive identification, which must again be unique, as we will see.
But if the second-order volition is not necessarily effective in determining one’s first-order will, what is important about second-order volition? Frankfurt focused on second-order volitions as a way to explain what it is for persons to identify with or alienate their own first-order states of will. A person who identifies with a desire in Frankfurt’s sense is not only conscious of the desire but regards it as his own in an additional or special sense: to be moved by that desire is in accord with the kind of character or person he wants to be. Identification is thus a reflexive inward state quite distinct from successfully determining one’s will. As opposed to a "wanton" who (like other animals) "does not care about his will," someone who is a person in the full sense has second-order volitions and is therefore not neutral or unconcerned about the will on which she acts: rather, she takes sides between her motives and identifies with one desire over others as the one she wants to be her first-order will.

This connection is first made in Frankfurt’s famous distinction between drug addicts. One is a "wanton" addict, who may both desire to take his drugs and desire to refrain from them, "cannot or does not care which of his conflicting first-order desires wins out." Another is an "unwilling addict" who "hates his addiction" and wants the desire to refrain from harming himself by taking drugs to be effective; thus when this healthy desire is overridden and he acts on the craving for the drugs, he is "helplessly violated by his own desires" (even though taking the drugs still counts as a human action or intentional movement, not a mere bit of behavior). This means that

The unwilling addict identifies himself...through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires.

---

627Ibid, p.16.
628Ibid, p.18.
629Ibid, p.17.
He makes one of them *more truly his own, and withdraws himself from the other.*\(^{630}\)

Thus on this account, it is by second-order volitions that we *identify* with a first-order motive, and in the process *alienate* other conflicting first-order desires on which we might possibly or do in fact act.\(^{631}\) ‘Identification’ in this intrasubjective sense gives the agent a personal or inward ‘identity’ apart from the economy of his first-order desires; by contrast, the wanton addict is not even really "neutral" towards his desires, since he has "no identity apart" from them:\(^{632}\) we might say that he lives wholly in the ebb and flow of passions, without concern about which of them win out and determine his will.

**IV.2.3. Identification, Freedom, and Rationality**

In the remainder of this paper, Frankfurt introduces two more themes that require mention. First, he uses his hierarchical distinction to ground a parallel distinction between *freedom of action*, which requires that the agent is "in a position to translate his first-order desires into action" or that she can perform the act she wants,\(^{1}\) and *freedom of will*, which requires that her second-order volitions effectively guide the first-order desire she acts on, or determine which desire becomes her will.\(^{633}\) Someone who is moved to act by the desire, they want,\(^{2}\) to act on may have freedom of the will in this sense while lacking

---

\(^{630}\)Ibid, p.18, my italics.

\(^{631}\)So understood, ‘alienation’ is simply the negative correlate of identification. The unwilling addict *alienates* the first-order desire — a craving for drugs — on which he nevertheless acts. That craving still becomes his will, which is to say that it determines the intention through which we would explain the addict’s behavior as a particular type of action (e.g. ‘getting high’ rather than ‘taking medication for an illness’). Thus his ‘getting high’ constitutes action with an intensional designation: it is not involuntary as sleepwalking is involuntary. But nevertheless, the very intention that makes this an act of ‘getting high’ is in a deeper sense against his inner will, or *alien* to him.

\(^{632}\)Ibid, p.18.

\(^{633}\)As Frankfurt’s account of moral responsibility makes clear, if the agent is a person, freedom of action also requires that the desire, on which she acts not be alien to her.

\(^{634}\)Ibid, p.20.
freedom of action, and for a wanton, freedom of the will is not even an issue.\textsuperscript{635}

Note that there is apparently a difference on this account between simply identifying with the desire\textsubscript{1} that also happens to move one in acting and having free will in acting on this desire\textsubscript{1}: a contingent or accidental correspondence between one’s volition\textsubscript{2} and desire\textsubscript{1} is sufficient for the former, whereas the latter also requires that the agent guides his will\textsubscript{1} or "secures\textsuperscript{636} the conformity of his will\textsubscript{1} to his second-order volitions," which is incompatible with "their coincidence" being merely a "happy chance"—as it would appear to be for the willing addict, for example.\textsuperscript{637} Also note, however, that even an agreement between one’s volition\textsubscript{2} and effective desires\textsubscript{1} is not actually necessary for identification, since we can identify with desires\textsubscript{1} that do not in fact move us in our actions (as the unwilling addict illustrates). This is important for distinguishing Frankfurt’s account from analyses by others. For example, given Gerald Dworkin’s initial conception of "authenticity" simply as "congruence" between the first- and second-order will, Dworkinian "authenticity" would be a stronger condition than Frankfurtian identification, but a weaker condition than Frankfurtian free will.\textsuperscript{638}

Second, Frankfurt argues that although "desires and volitions of a higher order than the second" are possible, an agent terminates this series non-arbitrarily when he "identifies

\textsuperscript{635}Ibid, p.19-21.
\textsuperscript{636}Ibid, p.20.
\textsuperscript{637}As Eleonore Stump says, for Frankfurt, an agent’s freedom of will requires an agreement between his second- and first-order volitions, and that "he has the first-order volitions he has because of his second-order volitions" ("Sanctification and Hardening of the Heart, p.213). The implications of this interesting and often overlooked difference between freedom of will and identification (which can be constituted by an accidental correspondence of volition; and will,) will be pursued in the projected additional chapter on free will.
himself *decisively* with one of his first-order desires.\textsuperscript{639} What such decisive identification involves is not made very clear here, but it represents a kind of *unity* in the higher-order will: "The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, and any higher level, remains to be asked."\textsuperscript{640} At this point Frankfurt appears to think of decisive identification as constituted simply by another structure of higher-order volition: the person has no third or yet higher-order desires conflicting with this second-order volition, and, *eo ipso*, no conflicting second-order desires about what he wants to be his will.\textsuperscript{641}

In his 1971 paper, Frankfurt is primarily concerned to distinguish the phenomena of identification from traditional capacities for practical judgment. Yet he admits that the formation of second-order volitions manifests a "capacity for reflective self-evaluation,"\textsuperscript{642} thus

In maintaining that the essence of being a person lies not in reason but in will, I am far from suggesting that a creature without reason may be a person. For it is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will\textsuperscript{[1]} and of forming volitions of the second-order.\textsuperscript{643}

However, though the capacity for reflective practical judgment may be necessary, it is never by itself sufficient for personhood, according to Frankfurt.

It is compatible with this position that there may be other necessary conditions implicit in the idea that a responsible agent can deliberate about her potential actions (though I do not commit myself to such conditions here). For example, Paul Benson has proposed another

\textsuperscript{639}Ibid, p.21.
\textsuperscript{640}Ibid, p.22.
\textsuperscript{641}I consider these relations between moral responsibility, identification, and "freedom of will" in the projected additional chapter on free will and responsibility for character.
\textsuperscript{642}Ibid, p.12.
\textsuperscript{643}Ibid, p.17.
explicitly interpersonal condition for autonomy and responsibility for one’s acts: in order even to trust her "possession and exercise" of the powers by which she regulates her actions and motives, a free agent must believe in "her own status as a worthy agent." He claims that this "sense of one’s selfhood" requires the belief that one is "competent to answer for one’s behavior in norm-governed relations with others," which in turn depends on others’ validation of one’s competence as an agent: if they treat the agent as insane, her freedom may be undermined even though "she may be able to commit herself to her effective desires through higher-order endorsement of them." She could, for example, be deceived into reflectively approving therapies to alter her beliefs and desires, but only because her sense that she is incompetent leaves her with no better option than to follow others’ better judgment. According to Benson, this "self-worth condition" shows that "unlike other sorts of intentional agent, persons are self-consciously social," and suggests that "freedom is valuable in virtue of being an element of persons’ self-respect as social beings."

Yet in drawing this quasi-Hegelian conclusion, Benson does not consider that agents could be social in the sense required by his self-worth condition without thereby necessarily acquiring the capacity for volitional identification. Even in combination, then, a sense of

---

646 Ibid, p.656. Benson considers examples of this deriving from the 1944 Bergmann film *Gaslight*.
647 Ibid, p.657. Benson takes this point to show that Frankfurt’s conditions for autonomy are inadequate, because the woman who has been ‘gaslighted’ “can satisfy any of the procedural conditions of freedom that have been proposed,” including structural or hierarchical ones: her unfreedom is not a failure of these conditions (p.657). Yet I think it is more plausible to take Benson’s example as evidence that a person’s higher-order volitions have to be *freely formed* to count as identifications constituting the inner character on which her moral worth depends: Benson’s condition shows that such free formation is incompatible with certain types of deception and manipulation which destroy the agent’s sense of self. Whether such free formation also requires *liberty* or alternate possibilities of identification will be analyzed in the projected additional chapter on free will.
ourselves as worthy agents, with its interpersonal conditions, and a normal human capacity for rational deliberation, are not together sufficient for volitional identification, because an agent could possess such "rational faculties of a higher-order" and believe itself competent before others in using these abilities, yet remain essentially wanton.\textsuperscript{649} This is because according to Frankfurt, identifications are never simply constituted by evaluations of one of our desires, as ethically superior or more rational than the other desires, on which we might act. Hence, though they may be necessary for persons, even the conjunction of such cognitive conditions as deliberative rationality, a capacity to understand one’s agency as minimally competent, and conscious individuality\textsuperscript{650} still does not capture what is distinctive of persons as such, since the capacity for volitional identification cannot be explained by the sum of conditions of this kind. It involves a further motivational capacity that is not reducible to cognitive capacities alone.\textsuperscript{651}

However, it remains tempting to assume, with Hume, that the only alternative is for the remaining condition of personhood to be purely appetitive. This may explain why in his "Freedom of the Will" paper Frankfurt says not only that "It may not be from the point of view of morality that the person evaluates his first-order desires" (which is obvious) but also that

a person may be capricious and irresponsible in forming his second-order volitions and give no serious consideration to what is at stake. Second-order volitions express evaluations only in the sense that they are preferences. There is no

\textsuperscript{649}Ibid, p.17.

\textsuperscript{650}This is the condition that most interested Strawson, of course.

\textsuperscript{651}In a later essay, Frankfurt argues that practical deliberation itself may depend on volitional capacities. Reasoning requires making up one’s mind and thus must have a hierarchy-structure involving in identification. If so, it appears that second-order volitions are necessary for reasoning: see "Identification and Wholeheartedness," in Responsibility, Character, and the Emotions: New Essays in Moral Psychology, ed. Schoeman; reprinted in The Importance of What We Care About: 159-176, p.176.
essential restriction on the kind of basis, if any, on which they are formed. Notice that this is a stronger claim than the phenomenological thesis that identification involves a motivational factor not constituted by the cognitive conditions philosophers have proposed for persons; in addition, it suggests that this motivational factor can be a brute desire or impulse. Thus although he started out using the term ‘desire’ only in a formal sense, Frankfurt at this point regards second-order volitions merely as what I called ‘subjective preferences’ or desires in the S2 sense (see Chap. II §2), making them compatible with a Humean psychology of motivation. Hence, while resisting rationalist reductions of identification to practical deliberation about our desires, Frankfurt briefly succumbed to the opposite danger of giving an emotivist explanation of identification, rather than articulating an alternative to both the rationalist and emotivist conceptions of motivation, as I tried to do in Chapter II.

IV.3. Objections to Frankfurt’s Initial Account of Identification

In sum, Frankfurt’s 1971 paper introduces the phenomena of volitional identification and decisive identification as central features of personhood, and in the process also gives his own positive account of these phenomena in terms of hierarchical relations between desire-states of certain kinds. Objections have arisen to both aspects of this contribution, though they have not always been clearly distinguished. Many of the objectors start in implicit agreement with Frankfurt that the phenomenon of volitional identification is somehow reflexive and genuinely distinctive of persons, because it is the driving force in ‘care of the self,’ or central to the experience of developing a ‘self’ that transcends the ebb and flow of

\footnote{Ibid, p.19, note 6.}
first-order desires. But these objections question whether Frankfurt’s *particular account* in terms of second-order desires adequately explains this phenomenon. Another class of objections is better regarded as rejecting the prior phenomenological claim that anything like the ‘identification’ Frankfurt describes is essential to personhood because it establishes who we are the deepest sense, or that volitional identification is the central factor determining the person’s ‘authentic self.’

**IV.3.1. Two Initial Objections Against Identification as a Central Phenomenon**

There are two main objections at this more fundamental level. The first concerns the question of why reflexive attitudes or higher-order volitions, rather than our impulsive desires or immediate appetites, should be more important to our ‘real self’ or be more important in determining our authentic identity. The second, as formulated by Thomas Pink, asks why persons as "decision-makers" need to make their motives for first-order acts into *objects of concern* at all when deciding or forming intentions to act.\(^{653}\)

**(Obj.A1) The ‘Animal-Self’ Objection.**

In her own list of objections to Frankfurt’s account of identification, Eleonore Stump presents the following concern:

Finally, some critics have objected to Frankfurt’s account on the grounds that (3) it rests on an unwarranted notion of what counts as "the real you" and on a false theory of what counts as external or alien to a person....Why should we identify the agent’s self only with his higher-order volitions. Why should we suppose that these higher-order volitions represent what the agent really wants? Has psychology not made us aware that the darker sides of our nature, the repudiated or repressed first-order desires, are just as much a part of our selves...?\(^{654}\)


I will refer this objection, which Stump derives from Irving Thalberg, as the ‘animal-self’ or Freudian objection. Its concern seems Freudian in inspiration, since it fears that claims for identification deceptively equate the self with something like the superego, rather than the id. But repression or failure to acknowledge our lower desires, animal appetites, or subconscious emotions will not lead to a well-integrated self. Indeed, on some versions of this view, our ‘real self’ is simply given, like a noumenal character or Leibnizian monad.655

The problem with this objection, as Stump rightly perceives, is that it fails "to distinguish two different senses of identification."656 She illustrates this with an imagined story of a woman who has a child by rape, and feels "conflicting attitudes towards the child," who looks like his father. She can either identify herself "as a hater of the child," thus assenting to her feeling of repulsion,657 or she can fight this feeling and identify with her affection for the child. Since she could do the former, "it is clear that an agent may identify herself with any of her first-order desires, no matter how savage or irrational they may in fact be."658 But even in such cases, in a different *constitutive* sense, her *self* still rests in her second-order volitions; her ‘person’ is identical with the part of her psyche that originates her identifications in the former sense (which for Stump must be her "reasoning faculty," as we

654(...continued)
211-234, pp. 218-19.
655Using these images, Carl Jung argued that the only route to personal "individuation" or self-cultivation was precisely *through* tapping into the archetypal subconscious or "collective" layer of the psyche, and being guided in one’s development by the images of personal wholeness arising out of it. See his books, *Psychological Types*, *Archetypes of the Collective Unconscious*, *Symbols of Transformation*, *Two Essays on Analytical Psychology*, and *Aion: Researches into the Phenomenology of the Self*.
will see).\textsuperscript{659} The fact that we can integrate or in some cases simply adopt elements of our ‘id’ as authentically our own thus \textit{depends on} volitional identification, rather than presenting any challenge to its central role in personhood.

Even when Nietzsche imagines the higher man as one who tries to let his animal instincts guide him, this is not something a \textit{mere} animal can do, for it implicitly involves \textit{identification} with the aggressive instincts Nietzsche celebrates in flouting Judeo-
Christian mores and traditional theories of virtue. Of course, Nietzsche insists on a purely Hobbesian account of such identification when he says: "The will to overcome an affect is ultimately only the will of another, or several other affects." Thus he would reject any account of identification as depending on motivational states qualitatively distinct from brute desires and inclinations. But he still recognizes the phenomenon of identification as such, even in embracing the ‘id.’

**(Obj.A2) The Davidson Objection.** Thomas Pink argues that "justifications for deciding to perform an action are desirable ends which taking that decision would further," and that reasons which justify a decision to A must also apply to doing A, giving the agent an instrumental reason for carrying out her intention to A: "desirable ends can only justify one’s taking a decision to act if, once that decision is taken, they would thereafter be furthered by, and so justify, one’s acting as decided." This means that a decision cannot be taken for merely strategic reasons unrelated to any motivation to do the act decided on. Pink uses this point in novel ways to undermine Gauthier’s recent claims that by extending rational decision theory to second-order acts, we can resolve the paradoxes of deterrence and give a purely ‘self-interested’ account of apparently altruistic dispositions. But this relatively weak connection between the motive to form an intention to A and the motive to A makes Pink sympathetic to Davidson’s claim that there is a much stronger connection:

---

660 See, for example, Friedrich Nietzsche, *Beyond Good and Evil*, tr. Walter Kaufmann (Vintage Books, 1989): "We think that hardness, forcefulness, slavery...the art of experiment and devilry of every kind, that everything evil, terrible, tyrannical in man, everything in him that is kin to beasts of prey and serpents, serves the enhancement of the species 'man' as much as its opposite does" (pp.54-55); and "'instinct' is of all kinds of intelligence that have been discovered so far, the most intelligent' (p.147)."

661 Ibid, p.86, Epigram 117.


663 Ibid, p.173-175.
In Davidson’s view, intentions are formed, and formed for reasons—explained by beliefs and desires which motivate and rationalize them as intentions— without being motivated and rationalized by beliefs about what one will be doing in forming those intentions to act. The beliefs which motivate and rationalize the intention are, rather, beliefs about what the agent will be doing in performing the action intended.  

In other words, though Pink disagrees with Davidson’s conclusion that decisions are therefore not themselves distinct actions, he agrees with his premise that

Practical judgments and decisions have a first-order content...Making judgments or decisions about how to act implies nothing more than a concern with what those judgments and decisions are about—a first-order concern with action. Second-order agency characteristically has a first-order motivation.

Pink does not appear to realize that this "nothing more than" claim is much stronger than his earlier proviso that the reasons for decisions must also apply to the actions decided on. Ironically, this stronger claim also seems to undermine Pink’s own point that by forming intentions to act in a given way (A), decisions perpetuate and strengthen the agent’s motivation to A: for if a decision to A itself is motivated only by the desire to A (or rationalized by its instrumental relation to the same end that rationalizes A), then how can it add anything to the motivation the agent already has in desiring the end to which A-ing is a means? In short, if decisions count as "acts" only because they can have an instrumental rationality based in the "means-end justifiability" of first-order acts, then decisions would seem to be motivationally epiphenomenal. Though Pink could not accept this implication, his allegiance to the Davidsonian emotivist dogma of rational decision theory leads Pink

---

664Ibid, p.43.
666Ibid, p.50-51.
667Ibid, p.60: "Again, standard decision theory is not ignoring any justifications for our desires when it assumes, as it does assume, that rational desires are simply a function of the likely desirability of their objects— and never a function of the likely desirability of holding the desires themselves." The problem with this, of (continued...)
to reject Frankfurt’s claim that

In making up his mind a person establishes preferences concerning the resolution of conflicts among his desires or beliefs. Someone who makes a decision thereby performs an action, but the performance is not of a simple act that merely implements a first-order desire. It essentially involves reflexivity, including desires and volitions of a higher-order. Thus, creatures who are incapable of this volitional reflexivity [essential wantons] necessarily lack the capacity to make up their minds. They may desire and think and act, but they cannot decide.668

Pink reads this as a claim that any sort of decision to act is itself a second-order act with "a motivation that is second-order"669 and protests that this misconstrues "second-order agency as we ordinarily conceive it," e.g. in cases such as deciding to see your dentist to cure a toothache.670 In this case, the first-order desire to be rid of the pain motivates going to the dentist, and the (prior) decision to go. By implying that your decision must instead be motivated by a second-order desire to be moved by the desire to ease the pain, Frankfurt’s account "foists onto decision-makers concerns which, simply as decision makers, they can perfectly well lack."671 Thus:

When you form a judgment that it is highly desirable to get rid of your tooth-ache, it is false that ipso facto you must care about your own motivation — that you must care about whether you desire to get rid of your tooth-ache...Again, if, on seeing you in danger, I decide to help...I need have absolutely no interest in whether or not

667(...)continued)


669Pink, The Psychology of Freedom, p.35.

670Ibid, p.36.

I am motivated to help you.\textsuperscript{672}

There are two strategies for replying to this Davidsonian objection. The first, simpler response is that, while many decisions or second-order acts may be motivated by nothing more than first-order desires, persons \textit{are nevertheless} capable of caring about which motives they generally act upon; thus they are also capable of acts of volitional identification in which the decision to form intention I1 rather than I2 depends on second-order motives \textit{beyond} the first-order desire for the ends embedded in I1.\textsuperscript{673} If you have been depressed and ignoring your health, you may decide to visit the dentist \textit{in part} because you identify with acting on desires that reflect your bodily needs, out of a determination to stop punishing yourself; alternatively, if you find yourself constantly running to doctors whenever you have the least little ache or pain, you may decide \textit{not} to visit the dentist right away in part because you explicitly alienate or do not identify with the disposition to act immediately on this type of impulse, since you don’t want to have a weak or hypochondriacal outward character. You need not have any such motives in making your decision about whether to visit the dentist, but as a \textit{person}, you can.

Moreover in many types of choice-circumstance, i.e. ones where our actions may have moral implications, this capacity for identification is important. It is plausible to read Kant’s famous division between "actions which in fact accord with duty" and action "from the motive of duty"\textsuperscript{674} as a distinction between actions simply motivated by some contingent

\textsuperscript{672}Ibid, p.38.

\textsuperscript{673}Frankfurt would add, of course, that we may nevertheless act on the contrary desire for end embedded in I2. Yet if we do, this may be \textit{in spite of} having formed I1. So in acting on the contrary desire, we may be acting without actively forming a deliberate intention to act this way at all.

\textsuperscript{674}Kant, \textit{Groundwork of the Metaphysics ofMorals}, Ak.397, p.65.
first-order desires whose maxims happen to be lawful or in accord with moral precepts, and actions whose maxim is adopted (whatever the desire, which supplies the end or "material" embedded in the maxim) because this maxim was lawful or furthered morally prescribed ends. In the latter case, there is a second-order motive for adopting an action-maxim which is not simply identical to the first-order motive embedded in the action-maxim (whatever it may be). Thus Kant would insist that in Pink’s second case, an agent of moral worth (rather than mere outward justice) does identify with her desire to help because she had a duty to aid in this circumstance.

Finally, Frankfurt need not be committed to the notion that every decision persons take need involve the deeper concerns of identification, with its existential implications. As he says, adult human agents who are persons rather than wantons may often "act wantonly, in response to first-order desires concerning which they have no volitions of the second-order, more or less frequently." These acts may involve decisions, but not the strongly evaluative commitments that Frankfurt has in mind in the passage for which Pink takes him to task (see §4.2 below). For some decisions do necessarily involve second-order motives:

---

675 Of course for Kant, these are still actions decided on, since the agent must form maxims to act, which will incorporate the ends of his first-order desires.

676 The fundamental problem with Pink’s whole analysis is his purely instrumental conception of practical rationality (see p.51), which dictates that the only motive deliberation can produce for doing something is that it serves as the best means to an end already within the agent’s ‘internal set.’ On this model, motives for deciding or intending to A that are distinct from direct motives to A (though they may still be such as to uphold the agent’s motivation to A once it is intended) because the kinds of motivation that typically satisfy these conditions are non-instrumental, e.g. the motive involved in identifying higher-order states is projective.


678 Pink is also wrong when he says that Frankfurt is trying to reconcile the Hobbesian idea that "There is no agency of intention-formation" (p.34) with the idea that freedom of will is something distinct from freedom of action by rejecting the other Hobbesian premise that all desire-formation is passive (p.33). It is false that for Frankfurt, "freedom of the will" only requires "desires with second-order content" and "does not require a new kind of psychological attitude —a state of intention which is quite distinct from any desire" (p.35). Frankfurt says explicitly that decisions create intentions to act, though the decision "does not guarantee that the intention will be carried out" ("Identification and Wholeheartedness," in The Importance of What We Care About, (continued...))
in particular, those by which a person resolves a conflict between opposing desires through identification must involve such motives. Pink protests that such conflicts can be "resolved at the first-order level" as long as the agent can form "beliefs that certain actions are desirable" and has "a strong enough first-order desire only to perform desirable actions." But, whether or not this is really a first-order desire, Frankfurt is primarily concerned about cases where such weakly evaluative beliefs themselves are in conflict. Moreover, since identification with first-order desires and motives "resolves" first-order motivational conflict only in the sense of determining the agent’s inward stand relative to it, but not necessarily in the sense of causing the motives the agent internalizes to be effective in guiding her actions, the decisions identification involves cannot depend for their motivation merely on whichever desire, is strongest: for the agent may reject or alienate even her strongest desire by identifying with a factually weaker motive. The phenomenon of identification is thus a counterexample to Davidson’s theory of motivation. Thus it begs the question to argue against the existence of identification as distinctive motivational phenomenon on the basis of Davidson’s theory.

A stronger reply is also possible. Pink not only ignores the importance of volitional identification and consequently ignores decisions that must involve distinct second-order motivation (e.g. the unwilling addict decides to visit the heroin pusher because she needs a high, but has still decided not to identify with this desire). He also fails to see that even in

---

678(...continued)
p.174).


680As Frankfurt himself notes, his examples show that, even though (as Davidson says) actions require intentionality and pertinent "attitudinal conditions" to motivate them, "the attitudinal conditions of a person’s action may themselves be alien to him," since whether he adopts them "as his own" is not determined by whether or not these attitudinal conditions eventually cause his action ("The Problem of Action," pp. 76-77).
the more prosaic decisions we make every day and hour —the apparently more ‘ordinary’ cases, such as deciding to see the dentist— there is often a kind of tacit identification with the relevant first-order desire in the decision. We let our decision to visit the dentist be motivated straightforwardly by our desire to cure the toothache in this situation because we have no particular qualms about acting on the motive of alleviating our pain, within reasonable boundaries. The fact that such a tacit identification is there in the background—or that the decision is not ‘wholly wanton’—is indicated by the fact that if the appropriateness of the first-order motives on which we decide is challenged, we can immediately take up the personal stance of an identifying agent, who explicitly considers adopting or alienating these first-order motives. A challenge that suggests that we have misjudged which desire\textsubscript{1} actually motivated our decision also automatically calls up this change in stance, because we may not be so tacitly comfortable with the alternative motive\textsubscript{1} that is suggested. Thus if our best friend says that we have decided to visit the dentist because we are in love with her, that heartache has more to do with it than toothache, we will ordinarily ask ourselves not only if this is the right diagnosis, but also, if it is right, whether such a motive is appropriate in our circumstance, or a good one on which to base our decision, or one that is in accord with the character we will for ourselves, etc. We are sensitive to such challenges because (whether we know it or not) we have defeasible assumptions in place that these challenges call into question, bringing into focus the tacit identification based on these assumptions, which was already given in the decision. Such tacit identifications have tacit second-order motives beyond the desire\textsubscript{1}, with which they tacitly identify the agent: otherwise, they would not be sensitive to challenges which do not actually diminish this desire\textsubscript{1}’s preferred position in our first-order field.

In conclusion, neither the Freudian nor Davidsonian objections give us any reason to
doubt the reality of volitional identification or its central importance in the formation of our ‘inward selves.’ Yet Pink protests against Frankfurt’s notion of second-order motivations in part because he follows Frankfurt’s early account in just equating such motivation with "second-order desire." He objects to the "fundamentally Hobbesian" idea that the will or second-order agency is simply "a form of desire," or that decisions in which we form intentions can be explained by "second-order versions of the very same attitudes" that explain or rationalize first-order acts. And this complaint may be justified. Although he does not distinguish identification from decision or intention-formation in general, and thus confuses Frankfurt’s analysis of it with a claim that we form intentions to form intentions, this aspect of Pink’s doubt is distinguishable from the Davidsonian objection concerning the validity of identification as a reflexive or intrasubjective phenomenon; it is more in line with several other objections that question the adequacy of Frankfurt’s particular analysis of the second-order motivation involved in identification as mere "second-order desire."

IV.3.2. Five Objections Against the Bare Hierarchical Account of Identification

Writers who raise these objections typically agree that Frankfurt has recognized something important for autonomous selfhood in describing identification and decisive identification, though they would revise his accounts of these central phenomena. As Gerald Dworkin says,

I think we fail to capture something important about human agents if we make our distinctions [to define autonomy] solely at the first level. We need to distinguish

---

681 Ibid, p.36.
not only between the person who is coerced and the person who acts, say, to obtain pleasure, but also between two agents who are coerced. One resents being motivated in this fashion, would not choose to enter situations in which threats are present. The other welcomes being motivated in this fashion, chooses (even pays) to be threatened.\footnote{Gerald Dworkin, "The Nature of Autonomy," in Dworkin, The Theory and Practice of Autonomy, p.17. Note that Frankfurt defines coercion to line up with diminished moral responsibility in such a way that the agent in the latter case would not count as "coerced."}

Although the level-distinction and reflexive character of identification seem to capture something indispensable, Frankfurt’s initial theory of how persons achieve the commitments of self which are necessary for autonomy has been criticized since its inception. Onora O’Neill rightly points out that (at this stage) Frankfurt seems to think of the will at every level solely as a disposition to satisfy preferences; thus even his second-order will can only be understood "instrumentally."\footnote{Onora O’Neill, Constructions of Reason (Cambridge University Press, 1989); p.66.} As a result, higher-order ‘volitions’ on this account are not qualitatively different in motivational structure from first-order desires, and although Frankfurt speaks of persons as "forming" their identifications, it seems that their second-order volitions could originate as passively as many of our first-order desires often do.

These limitations are responsible for five principal objections to Frankfurt’s first account, which appear in variant forms in the subsequent literature, which must be carefully distinguished. Stump summarizes the first of these objections as follows:

Various critics have charged that (1) on Frankfurt’s account, an agent would be free with respect to a certain volition\footnote{Stump. "Sanctification, Hardening of the Heart, and Frankfurt’s Concept of Free Will," pp. 218-19.} that was in accord with some higher-order volition even if that higher-order volition were directly produced somehow by someone else. But this is surely a counter-intuitive result, because such an agent seems to be just as much the puppet of his manipulator as the agent whose first-order volitions are directly produced by someone else...\footnote{Stump, "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will," pp. 218-19.}

This \textit{Origin of Identification} objection (\textbf{Obj.B1}), also includes Nagel’s concern that
"attempts to analyze agency in terms compatible with determinism — by reference to intentions, motives, second-order volitions, capacities,..." have made real advances, "but the possibility that these conditions are themselves determined seems still a threat to some element of the ordinary concept of action." Thus Obj.B1 really covers two closely related though distinguishable concerns: (i) the person’s liberty in forming higher-order volitions and (ii) any other limitations on the kind of process that leads to identificatory acts or dispositions. The first concerns what kind of alternative possibilities a person must have available in identifying, and the second asks what genetic conditions identifying states must meet to count as self-determined, both these questions concern the origination of identification in general, and they are closely related: as Watson notes,

Historically...the main inspiration of libertarianism is an understanding of self-determination that is incompatible with determinism; roughly, determinism means determination by something other than the self, and hence heteronomy. This construal of autonomy [or self-determination] thus dictates an indeterministic interpretation of alternative possibilities.

On the contrary, compatibilists like Frankfurt often argue that, far from entailing any indeterministic liberty of identification, self-determination may not only be compatible with, but even require strong limitations on our freedom to ‘identify otherwise.’ The issues of self-determination and liberty are thus inextricably linked. However, it is possible to ask (as in Obj.B1.ii) whether the process by which identifications originate is compatible with their

686 Thomas Nagel, “Subjective and Objective,” in Mortal Questions (Cambridge University Press, 1979), p.197. Nagel thinks this problem is inevitable, of course, since it results from the appeal of the "external view" to human reflection. But as I hope to show against Dennett in Chapter V, whether someone identifies with a motive cannot be defined in terms of criteria appealing to neutral or objective external judgments: it only makes sense from the inside.


688 Frankfurt’s most developed statement of the argument that autonomy conflicts with volitional liberty and requires some volitional necessities is given in "On the Necessity of Ideals;" but also see "Rationality and the Unthinkable" (in The Importance of What We Care About: 177-190) and Frankfurt’s 1991 Eastern Division Presidential Address, "The Faintest Passion," Proceedings and Addresses of the APA, Vol. 66 (1992): 5-16.
*self-determination*, without expecting this process to involve libertarian choice. As Watson puts it,

> [this] problem concerns the source of one’s higher-order volitions. For all [Frankfurt’s initial] account tells us, the person’s higher-order preference may be the result of brainwashing, or severe conditioning of the kind which is plainly incompatible with autonomy. This is of course a standard problem for compatibilism and the hierarchical version only moves it up a level.\(^{689}\)

John Fischer has particularly stressed this as a problem for Frankfurt’s view that persons are morally responsible when they act on motives with which they identify, whether or not they could have done otherwise (e.g. like the "willing addict\(^ {690}\)). For if someone’s identification is constituted simply by a desire\(_2\) to act on a desire\(_1\), then

> his higher-order volitions and ‘identifications’ could also be directly induced. But this sort of agent does not act freely (insofar as acting freely suffices for moral responsibility). He is not a rational candidate for the reactive attitudes.\(^ {691}\)

In other words, Fischer takes it as obvious that if identification-constituting states are brought about by neurological interference, conditioning, hypnosis, or some other direct form of manipulation, the agent would not be responsible. If we say that identification cannot originate in responsibility-defeating ways, the account of it must be historical, not merely structural; if we say that is can originate by manipulation, it will divorced from moral responsibility.\(^ {692}\) Fischer has added that any "hierarchical mesh theory" focusing on "selected arrangements of elements internal to an agent’s motivational systems" will be subject to the objection that its conditions for identification could be satisfied in "a responsibility-undermining way —say, by direct electronic stimulation of the brain,

---


\(^{690}\)Frankfurt, "Freedom of the Will and the Concept of a Person," p.24-25.


\(^{692}\)Ibid, pp.50-51.
subliminal advertising, and so forth.\footnote{Fischer, \textit{The Metaphysics of Free Will}, p.208. I return to this issue in the projected chapter on free will.}

Yet this objection (B1.ii) may be open to various compatibilist avenues of response: for example, that the process by which identifications are formed would have to be in some way counterfactually "reasons-responsive," as Fischer\footnote{See Fischer, "Responsibility and Control," and \textit{The Metaphysics of Free Will}, ch.8.} and Christman\footnote{John Christman, "A Defense of the Split-Level Self," \textit{The Southern Journal of Philosophy}, Vol. 25, No.3 (1987): 281-293, esp. 289-291.} have suggested, or "procedurally independent" of "hypnotic suggestion, manipulation, coercive persuasion, subliminal influence" and the like, as Gerald Dworkin argues.\footnote{See Dworkin, \textit{The Theory and Practice of Autonomy}, p.18.} Since such conditions might be satisfied by agents whose identifications were still causally determined, but by the right sort of causes, Obj. B1.ii does not pose the same challenge as Obj. B1.i, since the latter protests that identification must involve choices made with libertarian freedom.

It is important to recognize that both concerns comprised under Obj.B1 are concerns about whether the picture of identification given on Frankfurt’s initial model can fit with our considered intuitions about conditions for \textit{moral responsibility}. The remaining objections differ in kind from Obj.B1, because they are not about the adequacy of Frankfurt’s model of identification for the theory of moral responsibility; rather they concern \textit{the direct phenomenological adequacy} of Frankfurt’s model for our experiences of identification themselves (the theme on which this chapter focuses).

The familiar "regress problem" for hierarchical accounts of identification is an objection to the ability of the desire-model to explain the phenomena of identification in this sense. Stump summarizes this objection as follows:

Another recurrent complaint is that (2) Frankfurt’s account of free will leaves us with an infinite regress of volitions. For an agent to act with free will with respect...
to some first-order volition V1 apparently requires that he have a higher-order volition V2 with which V1 is in accord. But then it seems that V2 itself must be freely willed in order for the agent to be acting with a free will; for V2 to be freely willed, however, requires a higher order volition V3 with which V2 is in accord...

Although Stump phrased this objection as a problem about the conditions for free will, it really expresses an underlying problem with the account of identification as a *hierarchical relation of desires*. In 1975, Gary Watson argued that Frankfurt’s hierarchical distinction cannot give an account of the sense in which some wants may be said to be more truly the agent’s own than others (though in an obvious sense all are wants of the agent), the sense in which the agent ‘identifies’ with one desire rather than another and the sense in which an agent may be unfree with respect to his own ‘will’.

Because it does not tell us why one can’t be "wanton, so to speak, with respect to one’s second-order desires and volitions." As Watson restated it in 1987, the problem is that the "Hobbesian account" of identification in terms of mere desires overlooks the fact "that desire can itself be seen as impedimental" to autonomy, and since higher-order volitions on this account are simply desires, "nothing about their level gives them any special authority."

Similarly, Robert Nozick says that

An addict who desires not to desire heroin may know that he cannot feasibly obliterate his first-order desire for heroin, and thus know that the only way to resolve the conflict of preferences is to drop his second-order desire not to have that first-order desire.

______________________________


Why, Nozick implies, should he necessarily rate his second-order desire are more important? Dworkin also mentions this problem, and notes that even procedural conditions on the origin of identification (that might solve Obj.B1.ii) will not resolve it, since not being manipulated into altering our second-order desire to fit our actually dominant first-order desire does not seem sufficient to make us "autonomous" as a result.\(^{702}\)

This counterintuitive possibility of becoming autonomous by ‘selling out’ to oneself, or letting one’s strongest desires\(_1\) guide one’s volitions\(_2\) and thus getting agreement between them ‘on the cheap,’ is a symptom of the fact that analyzing volitions\(_2\) as desires\(_2\) robs them of any higher authority than one’s first-order desires.\(^{703}\) We can see that this is directly related to Watson’s concern, because if autonomy by ‘selling out’ were possible, then the wanton would only need to add a will\(_2\) to approve whatever desires\(_1\) happen to move him at the moment, no matter what they might be, to become autonomous. But such a mere ‘rubber stamp’ approval would be superfluous, and could hardly make him less “wanton:” rather, his

---

\(^{701}\) Though Nozick himself does not present it as a criticism: he uses it only in discussing a proposed rule that rational persons should have a third-order preference not to let conflicts between second- and first-order preferences stand. This analysis is foreign to the issue of identification.

\(^{702}\) Dworkin, *The Theory and Practice of Autonomy*, p.16: "Suppose that there is a conflict between one’s second-order desires and one’s first-order desires. Say one is envious but does not want to be an envious person. One way of becoming autonomous is by ceasing to be motivated by envy. But another way, on the view being considered here, is to change one’s objections to envy, to change one’s second-order preferences. Now there may be certain limits on the ways this can be done that are spelled out in other necessary conditions which I elaborated: that of procedural independence. So, for example, it wouldn’t do to have oneself hypnotized into identifying with one’s envious motivations. But even if the procedures used were ‘legitimate,’ there seems to be something wrong with the idea that one becomes more autonomous by changing one’s higher-order preferences” to fit any pre-existing first-order motives.

\(^{703}\) Dworkin does not seem to recognize this, however, since (1) his new model of autonomy still understands identification in terms of “higher-order preferences” (p.20) and (2) he treats this problem of autonomy by ‘selling out’ as quite distinct from the regress and ab initio problems (p.19). Also note that Frankfurt’s stronger condition on free will, which differentiates it from Dworkin’s “authenticity” as we saw, seems to anticipate this problem. But since having free will or having one’s identifications be effective is not essential to identification per se on Frankfurt’s initial model, the stronger conditions on free will cannot help with the ‘sell out’ problem.
wantonness would then simply infect his higher-order desires.\textsuperscript{704} We are not, then, automatically identified with second-order desires in virtue of their being about other desires: identification with them is as contingent as identification with other states. Thus as Stump says, citing Watson,

If an agent is wanton with regard to his first-order desires unless he has second-order desires with which he identifies, why shouldn’t we suppose that he is wanton with respect to his second-order desires unless he has third-order desires with which he identifies. But then we would also need fourth-order desires to ward off wantonness with regard to third-order desires, and so on.\textsuperscript{705}

This formulation combines the regress issue with the issue of parity-of-authority between first and second-order desires.\textsuperscript{706} Jan Bransen has helpfully divided the problem into two parts:

\textbf{[Obj.B2]} The first is the threat of an infinite regress: if a first-order desire is autonomous because the agent has a second-order desire that this first-order desire that this first-order desire be effective, then the agent seems to be in need of a third-order desire to account for the autonomy of this second-order desire, and a fourth-order desire...\textit{ad infinitum}...

\textbf{[Obj.B3]} The other problem is the so-called \textit{ab initio} problem: how can a higher-order desire guarantee that an agent acts autonomously in following a lower-order desire if we do not seem to need a guarantee that the agent acts autonomously in following the higher-order desire? In short, why should we go to the second-level at all?\textsuperscript{707}

\textsuperscript{704}It is interesting, given the origins of the concept of identification in the Stoic tradition, that one perverse mutation of Stoic eudaimonism tends to assume that precisely this sort of approval of whatever condition we are stuck with will make us happy and fulfilled. Nietzsche, who inherits this idea, proclaims in his doctrine of the \textit{amor fati} that joyous approval and celebration of whatever happens (both in the world, and in our motivations), to the extent that we could will its infinite repetition, is the mark of the freed man, the \textit{übermensch}.


\textsuperscript{706}Though Stump notes, without citing Bransen, that "the problem of authority and the problem of infinite regress are two sides of the same problem" (ibid, p.185).

These two objections together form a dilemma for Frankfurt’s first model of identification: if a desire (of any order $n$) is autonomous only because we have another desire (of order $n+1$) for the first desire, then the application of this condition to the higher-order desire itself generates an infinite regress; but if this condition does not apply to the desire$_{n+1}$, then why should it apply to the desire$_n$? (why can’t we say that a wanton’s first-order desire alone is sufficient for her identifying with it?). As Bransen aptly notes, this dilemma forces apart the intuitive connection Frankfurt draws between two senses of ‘hierarchy’ or ‘high-orderness:’ in the weak or thin sense, "That a mental state is higher means [only] that it is about a lower state, or has this state as its intentional object;" but in the strong or thick sense, "That a mental state is higher means that it articulates with more authority the content of a person’s will," or carries identification with her real self.\textsuperscript{708} As a result of Obj.B2 and B3, however, these senses come apart:

Although it is clear that the hierarchical account of autonomy derives its plausibility from the assumption that both connotations are related, it is not obvious how or that they are related...if it is clear that a desire is not automatically more authoritative about a person’s will merely by being higher up in the hierarchy, then it is clear that the account still lacks an explanation of the way in which such a desire becomes authoritative.\textsuperscript{709}

Objs.B2 and B3 thus pose a dilemma for hierarchical account of identification. But it is important, as Bransen notes, that the threat of these objections "concerns the account of autonomy, not the experience of autonomy."\textsuperscript{710} The problem is about how to conceive identification, not about whether we experience this phenomenon. To avoid this dilemma, we need a hierarchical conception in which the conditions for autonomy differ intrinsically

\textsuperscript{708} Ibid, p.2 (the weak/thin and strong/thick descriptors are mine, not Bransen’s).

\textsuperscript{709} Ibid, p.2.

\textsuperscript{710} Ibid, p.14, note. 10.
between the motivational states of different orders whose relation is used to explain identification. But there does not seem to be such a difference with desires in the S2 sense. For example, economists are familiar with situations in rational decision-making where an individual apparently has desires about his own preferences that help determine which of them is most prudent to pursue in different circumstances. For example, if I have the (S2-type) desire to maximize the satisfaction of my inclinations, or preferences, the fact the first desire is weakly higher in the sense that it is about other desires (or that an iteration of "desire" states is the easiest way to describe this complex preference does not mean that it has any intrinsic internality for me, or is any more authoritative for my will than a simple addictive desire, with which I can identify or not. Indeed, I could view this entire complex desire for utility-maximization as something repulsive, or something I’m determined to overcome. It could be as alien to me as the unwilling addict’s dominant impulses are to him. Nor would being governed in my actions by this instrumentally rational second-iteration preference ensure that I am not wanton with respect to my own identity.

If this case seems too abstract, consider the following one instead. My wife wants to see a movie. I also experience a lingering desire to see a good comedy film on television: call this S2-desire or subjective preference P1. But as it happens, I’m not much moved by P1 at the moment: I’m lethargic and feel like channel-flicking instead. Call this desire for mindless distraction P2. Yet I also experience an habitual urge to please my wife. This is an open-ended second-order S2-desire to act on motives in accord with her preferences. In the present circumstances, it is specified as a desire to act on P1. But this does not mean that I identify with P1. On the contrary, I might inwardly be determined to get work done, and

\footnote{For in fact, as we saw in Chapter II §3 (5), such cases will usually not involve genuinely independent second-iteration desires.}
actively identify with the desire to press on with my writing. In other words, I might view P1, P2, and P3 as unwelcome addictions, and regard their conflict as a no-win situation relative to my real commitment. Thus although P3 lends its force to P1, helping it win out against P2, I will to resist the motivation of which P3 is but a part. This P3→P1 complex as a whole remains first-order in the thick sense of its relation to my agent-authority, although P3 is about other motives.

By themselves, then, mere iterations of "desire" in the non-evaluative or merely preferential sense (S2) need not indicate anything about the sort of person one authentically wants to be. Thus it is not possible to give an account of identification that squares with Humean or emotivist moral psychology. This has dramatic implications: the structure of identification cannot be explained, for example, on the kind of moral psychology that Bernard Williams defends in "Internal and External Reasons" and later papers, or that which underlies classical economics and contemporary rational decision theory. Thus if identification is a genuine intrasubjective phenomenon, then these moral psychologies are clearly falsified.

Instead, to account for identification, our moral psychology must acknowledge motivational states that are qualitatively distinct from desire in the S2 sense. Identification may be a state that is ‘higher-order’ in the weak sense of being about first-order motives (or

See Williams, "Internal and External Reasons," reprinted in Moral Luck. My point here supplements my brief remarks on this paper in Chapter III §2. I do not mean that the phenomenon of identification shows that "external reason statements" (Moral Luck, p.106) or rather someone’s belief in such statements (p.107) could explain his action by themselves. Williams may well be right that purely cognitive states such as beliefs cannot by themselves motivate unless a desire or emotional disposition to act is already necessary to have such beliefs (p.107-8). Rather, what identification challenges is Williams’ implicit assumption that an agent’s “internal set” consists entirely of desiderative motivational states.

a dispositional pattern of them), but its special authority or ‘higher status’ in the strong sense will still have to be explained (to whatever extent it can be) by the qualitative difference between identification and S2-desire, rather than merely by its being about first-order states. In particular, to resolve the regress paradox, any theory of autonomy must recognize that the states which constitute identification (to whatever extent they can be isolated) are inherently or ‘always already’ identified with the agent, without needing confirmation by any further independent state to make them representative of the agent’s real self, or inward identity. In short,

(1) If $M_1$ is a motivational state with which (by itself) the agent is not necessarily identified, that is, a state that is in itself identification-neutral or wanton,\(^{714}\)

(2) And if a state $S$ carries or conveys agent A’s identification to $M_1$;

(3) Then $S$ itself is immediately and essentially identified with A, or has an authority that is not conveyed to $S$ by any other state: hence A cannot be wanton with respect to $S$.

This is the Qualitative Condition for identification: it requires that a person’s identification with the qualitatively distinct states constituting identification with other states is essential, as opposed to identification with first-order desires, emotions, and dispositions to act, which is never more than contingent. Another way to phrase this condition is as follows: whatever states or relations convey identification to other states are themselves identified with the agent without this identification being conveyed to them. In this difference between conveyed and contingent vs. non-conveyed or essential identification lies the basic disanalogy needed to block the dilemma formed by the regress and ab initio objections. This explains why, as we will see in Chapter V, most authors who recognize the regress and ab initio objections, but also recognize that identification is a genuine

\(^{714}\)In several essays, Frankfurt refers to such states at the “raw material” of psychic life, the primary ingredients of character.
phenomenon integral to the experience of moral selfhood, have offered accounts of it diverging significantly from Frankfurt’s initial model and conforming to this Qualitative Condition.

In "Identification and Wholeheartedness," Frankfurt explicitly recognizes Obj. B2-B3 in the forms posed by Gary Watson. He explicitly distinguishes "a lack of coherence within the realm of the person’s higher-order volitions themselves," which means a conflict or ambiguity of cares, from conflicts between the agent’s higher-order volitions and "the first order desire that is most effective in moving him." In the latter case (unfree will), an identified volition opposes an alienated desire, whereas in the former case (ambiguity), two identifications or states of care themselves are opposed. Moreover, the person with an ambiguous higher-order will is implicitly distinguished from the wanton agent because the former lacks "wholeheartedness" in his identifications, rather than simply lacking them altogether:

The second kind of conflict is within the volitional complex. In the absence of wholeheartedness, the person is not merely in conflict with forces "outside" him; rather, he himself is divided.

But although phenomenologically the difference is clear, Frankfurt concedes that his initial hierarchy of desires model cannot explain why an individual (especially a ‘halfhearted’ one) is "less wanton" with respect to her volitions than the wanton is with respect to her desires.

This phenomenon of ‘volitional ambiguity’ is relevant to Obj.B2-B3 in the following

---

717 Ibid, p.165.
718 Ibid, p.165.
way. The possibility of ambiguity re-emphasizes that its *uniqueness* or greatest strength among second-order desires cannot be what endows a higher-order volition with authority, since a conflict among identifications does not make the agent as passive as a wanton. But if identification *was* constituted merely by desires$_2$ or preferences$_2$ about what motive$_1$ to act on, then volitional ambiguity would be a conflict among desires$_{s_2}$ which looks like an *exact parallel* at the second-order of the first-order "economy of passions" in which the wanton immerses himself. This is a *reductio* of the initial model, because it makes intuitively clear that "The mere fact that one desire occupies a higher level than another in the hierarchy" is "insufficient to endow it with greater authority or with any constitutive legitimacy"$^{719}$ (the *ab initio* problem). To identify us with first-order states, volitions, themselves must be ones "by which the person really wanted to be determined" (i.e. they must be identified with the self) and this cannot be because of yet-higher order desires (the regress problem).$^{720}$ Thus Frankfurt acknowledges the inadequacy of his initial model.

Beyond these objections, however, there still remain two other separate problems for the initial model. The first, which I call the *Sporadicity Objection* (Obj.B4) concerns the apparently instantaneous character identification: as Frankfurt first construed it, we could be "capricious" in forming second-order volitions,$^{721}$ and thus it seems we could identify with our desire$_1$ at one moment and identify with an opposite desire$_1$ at the next, without ceasing to be autonomous. As Benson says, on "structural" versions of "procedurally defined conceptions of freedom" such as Frankfurt's, agents "regulate their wills just when the motives$_{s_1}$ that lead them to act stand in a certain relation to certain other *contemporaneous* ...

$^{719}$Ibid, p.166.
$^{720}$Ibid, p.166.
states\(^{[2]}\) of the agent.\(^{722}\) But as Dworkin says, autonomy is not best conceived as a property of very discretely individuated acts (e.g. picking up a toothbrush) but rather as a broader feature of one’s character:

It...evaluates a whole way of living one’s life and can only be assessed over extended periods of a person’s life, whereas identification [on Frankfurt’s and Dworkin’s initial accounts] is something that may be pinpointed over short periods of time. We can think of a person who today identifies with, say, his addiction, but tomorrow feels it as alien, and who continues to shift back and forth at frequent intervals. Does he shift back and forth from autonomy to nonautonomy?\(^{723}\)

The issue here concerns what I called the proper object of identification (§1 above). But note that Dworkin, somewhat confusingly, continues to use "identification" to stand for the time-slice relation of harmony in Frankfurt’s initial account, and uses "autonomy" for a state that involves more continuity than ‘identification’ in this inadequate synchronic sense. But the force of his objection is really to point out that a desire, emotion, or any motive,\(_1\) for acting cannot really be ‘the agent’s own’ in the deeper or authoritative sense if it is so arbitrarily and reversibly preferred or not preferred. In our terms, the phenomenon of identification as we experience it involves more persistence of inner alignment with our motives,\(_1\).

As we will see in more detail, qualitative reinterpretations of identification must take account of this inertia in identification: this is part of what qualitatively distinguishes it from mere preferences which, given their lack of objective evaluational content, can be in almost

---

\(^{722}\)Paul Benson, "Free Agency and Self-Worth," p.653. He continues: "For instance, Frankfurt claims that a free agent’s identification with the motives upon which he acts consists in the agent’s acting upon those motives (partly) because they are endorsed by the agent’s highest-order motives concerning the situation." While this exposition wrongly adds into identification a feature that is instead part of free will for Frankfurt —namely that one’s second-order volitions effectively guide one’s will,— it is correct that the initial model of identification takes an atemporal time-slice view of the agent.

\(^{723}\)Gerald Dworkin, *The Theory and Practice of Autonomy*, p.16.
constant flux. In fact (as I suggested in IV §1) our actual identifications seem more like settled *dispositions* to identify with *types* of motives, e.g. desires to maintain good bodily hygiene, rather than point-like identifications with this particular desire (which has just such-and-such a cadence or ‘feel’) to brush my teeth now. The qualitative revision will thus both clarify the central role of identification in forming our entire motivational *character*, and help us see how we can be said to identify with particular desires even when we don’t have time to reflect on them before acting on them.

The last remaining major objection to Frankfurt’s initial account instead concerns his conditions for *decisive identification*, since it is closely related to the fact that higher-order volitions themselves may be in conflict. Following Gary Watson’s formulation of it, I will call it the *Arbitrariness Objection* (Obj.B5): an agent cannot decisively commit herself to acting on some motive, simply by having a volition to act on this motive, and refusing to engage in any further reflective evaluation of this volition, which could potentially lead to conflicting yet-higher-order volitions, because on this account, the agent just arbitrarily stops the regress. As a result, neither could the agent decisively identify with one of two (or more) conflicting volitions, simply by forming another desire that one of them be effective, and leaving this desire undisturbed by any other desires of the third order or higher.

As I will argue (see §5.1 below), answering this objection (B5) should thus not be confused with answering the dilemma involved in the regress and *ab initio* objections, which points towards the need for a qualitative account of identification. Rather, the problem of

---

724 At one moment, I may think I prefer chocolate ice cream, but as soon as I taste it, wish I had gotten vanilla instead. Yet the next day, the chocolate may seem better to me again. Modelling my volatile changes in terms of preferences is no problem because preference-hierarchies may constantly be adjusted in light of new information, or even for no particular reason.

decisive identification properly understood *presupposes* a qualitative revision that would resolve the prior problem. If we can account for positive identifications (or identifications *per se*) as constituted in some qualitatively distinct manner, we still have the problem that, so constituted, *they* can conflict, and so we need a higher notion of identification that can overcome and rule out such conflicts. When faced with this problem, we will not be able to resolve it by appealing simply to third-order desires, or to the agent’s simply *preventing* any conflict with one of his identifications in any arbitrary fashion. Thus an explanation of how this higher or decisive sort of identification is constituted cannot resolve the *prior* question about how the level of identification whose ambivalences it resolves itself first gets constituted.

For this reason, I will leave further analysis of the implications of Obj.B5 until later. But we can already see that this problem implies that, just as deliberate identification is qualitatively distinct from a wanton’s *consciousness* of her motives, for acting as informing *her* intentions (having motives, does not *ipso facto* mean positively identifying with them), decisive identification must also be qualitatively distinct from positive identification (since identifying with a motive, does not *ipso facto* mean decisively identifying with it or —which is the same thing— with the identification itself). Although states constituting identification must themselves *eo ipso* already be identified with the agent in their very constitution, according to the Qualitative Condition, this does not mean that they necessarily carry *decisive* identification with them as well. Accordingly, we should recognize that there are different levels or *orders* of identification itself in the strong sense of ‘order:’\(^\text{726}\) decisive or higher-order identification is more authoritative or reflects the agent’s inner self in a

\(^{726}\)See IV §6.1 for a fuller explanation of the three different orders.
distinctively deeper way. Within identification as a broad intrapersonal set of phenomena, there are further *qualitative* differences of order which do not supervene on nonqualitative ordinal differences in the iteration of desires.

### IV.4. Developments in Frankfurt’s Later Accounts of Identification

The previous section has made clear that Frankfurt’s initial model of identification was too ‘emotivist.’ While Frankfurt distinguished identification from the operations of rational reflection and conscious intentionality which were given pride of place in preceding theories of personhood, he also rejected *any* restrictions on how identifications could be formed or grounded (as we saw at the end of §2), and thus fell into the opposite fallacy of holding that identifications can simply consist in ordinal relations among weakly evaluative preferences or desires in the non-cognitive (S2) sense. For some, this gave his account extra appeal: the idea that phenomena of identification could be resolved into complexes of desires implied that the distinctive feature of personhood could easily be accommodated within a naturalistic ontology.

Nevertheless, while Frankfurt certainly seems sympathetic with the dominant view in recent Anglo-American moral theory that ethics should be pursued with the least amount of metaphysical baggage possible, later developments have shown that it was never his intention to replace a Strawsonian conception of personhood with a purely emotivist conception instead. Frankfurt rejects Hume’s extreme view that desires cannot be *irrational* in any normative sense, and so he cannot hold that personhood is reducible to a set of ‘passions’ to which *reason* would be a mere "slave." Moreover, in later papers, Frankfurt

---

727Frankfurt, "Rationality and the Unthinkable," new in *The Importance of What We Care About*: 177-190, esp. §IV p.184-188.
has developed ideas leading towards a qualitative account of identification.


The first indications of this shift appear in Frankfurt’s papers on "coercion" as a condition excusing the agent of moral responsibility for his action. Frankfurt argues that this condition cannot adequately be defined without the notion of identification: coercion involves an offer or threat that "appeals to desires or motives which are beyond the victim’s ability to control" where the victim also does not identify with acting on these motives. For a person may act on an irresistible motive, aroused by an offer or threat, while

the desire which finally motivates him when he performs the action is one by which he in no way resents being motivated. He has no desire or inclination to resist the desire which moves him into compliance with the terms of the offer, nor does he in any way regret being motivated by it.

Surely this person is not coerced....The fact that the desire which moves him is irresistible is also consistent with his autonomy, since he identifies himself wholeheartedly with this desire. Its thrust —though in fact beyond his capacity to control— in no way diverts him from the pursuit of his own aims.

This formulation already suggests that there is something more to identification than mere preference, for a person who prefers A to B to C to... in a simple ordering may, on finding that A is unavailable, go with B without much resentment or regret, if she considers her second-best option not much worse than her first: settling for B instead of A may not in any deep way frustrate her long-term aims or life-goals, even if A, B, C etc. refer to desires on which she acts. Yet Frankfurt also suggests that threats are unlike offers in that "every effective threat" violates the victim’s autonomy, because "in submitting to a threat, a person invariably does something which he does not really want to do," even if he identifies with

---

729 Ibid, p.41.
other desires that would motivate actions similar to those which the threat motivates. This is because a person who submits to (rather than merely complying with) a threat "necessarily does so in order to avoid a penalty...his motive is not to improve his condition but to keep it from becoming worse. This seems sufficient to account for the fact that he would prefer to have a different motive for acting." Such an explanation still reduces identification to a mere preference for one desire rather than another: the agent fails to identify with the threat-motive (TM), not because of any evaluation of the intrinsic qualities of TM itself —such as that fear is ‘unmanly or a dehumanizing or debasing motive— but simply because he would prefer not to be in the kind of situation S that TM presupposes. If this were right, then the agent’s alienation of TM would be derivative from his desire to enjoy more pleasures or satisfy more of his preferences: S precludes such an ‘improvement’ so, he prefers not to act on any desire (such as TM) that presupposes S. But if this analysis held, then alienation and identification would indeed be explainable purely in terms of first-order motives, as Pink’s objection (Obj.A2) proposed. Such an account is phenomenologically inadequate, because it does not reflect the fact that unlike the motives involved in identification, all such first-order motives are themselves alienable: none of them satisfies the Qualitative Condition.

In a subsequent essay on coercion, Frankfurt illustrates a similar ambiguity in responding to Don Locke’s objection that a "Devil/neurologist" (D/n) could ensure that

---

730Ibid, p.42-43. Frankfurt’s example is a man who identifies with his benevolent desire to give his money to the first person he meets, and is then held up by a robber threatening him with a pistol. Although handing over the cash is outwardly similar to the action of giving the money, it is done with a different intention (Frankfurt says it is the same action, but we must remember that for him this refers to the intention that guides the bodily movement, i.e. the intent to transfer money by moving one’s arm, not the deliberate intention to guide one’s movements this way for a purpose or end).

731Ibid, p.44.

agents act freely or without excusing coercion on Frankfurt’s model—a version of Objection B1 as outlined above. Frankfurt concedes that a subject whose mental and physical states are all the product of constant manipulation by a devil or neurologist could have "second-order desires and volitions" without having any "character or dispositions of his own," or any that constitute a self. Such a subject would lack "even the minimum of continuity and intelligibility essential to being a person," since the "identifiable themes" in his first-order willing would not be "internally rooted," i.e. they would only have meaning from a third-person point of view. Thus, such a subject would lack the "inherent connectedness" of a person: his second-order volitions would not constitute identification with first-order desires among his mental states, and he would remain a wanton. This is tantamount to admitting that a desire$_2$ to act on some desire$_1$ that is simply inserted by an external manipulator and thus has no basis in the agent’s lasting motivational character lacks the quality necessary to convey identification, or to count as coming from the agent’s inner self.

Nevertheless —perhaps with Leibniz in mind— Frankfurt argues that this subject could become a person with moral responsibility for his action if the (D/n) "provides his subject with a stable character or program" in which his desires thereafter flow from this program and are thus continuously ordered. Then, given that the subject’s program includes certain second-order desires, he may become morally responsible, assuming he is suitably programmed, in the same way as others do: by identifying himself with some of his own second-order desires, so that they are not merely desires that he happens to have or to find within himself, but desires that he adopts or puts himself behind. In virtue of a person’s

---


734Ibid, p.53.
identification of himself with one of his own second-order desires, that desire becomes a second-order volition.\textsuperscript{735}

This again recognizes the qualitative distinction between higher-order volitions relevant for moral responsibility and mere higher-order desires. But the question remains: could the D/n have programmed his very identification with these desires? Frankfurt’s argument that the D/n could indeed "cause[] his subject to have and to identify with certain second-order desires" without affecting "the moral significance of the subjects’s acquisition of the second-order volitions" implicitly appeals to the very Qualitative Condition on identification I outlined above:

a person is active with respect to his own desires when he identifies himself with them, and he is active with respect to what he does when what he does is the outcome of his identification of himself with the desire that moves him in doing so. Without such identification the person is a passive bystander to his desires and to what he does, regardless of [how they originated]....As for a person’s second-order volitions themselves, it is impossible for him to be a passive bystander to them. They constitute his activity —i.e. his being active rather than passive— and the question of whether or not he identifies himself with them cannot arise. It makes no sense to ask whether someone identifies himself with his identification of himself, unless this is intended simply as asking whether his identification is wholehearted or complete.

The notion of identification is admittedly a bit mystifying, and I am uncertain how to go about in explicating it. In my opinion, however, it grasps something quite fundamental in our lives, and it merits a central role in the phenomenology and philosophy of human mentality. Instead of attempting to provide the analysis the notion requires, I shall limit myself to a declaration: to the extent that a person identifies himself with the springs of his actions [his motives\textsubscript{1}], he takes responsibility for those actions and acquires moral responsibility for them...\textsuperscript{736}

And since identification is built into the states that constitute identification with a motive, we automatically take responsibility for our identification-conveying volitions themselves, however they are caused.\textsuperscript{737}

There are several things to notice about this important passage. First, though it explicitly

\textsuperscript{735}Ibid, p.53.
\textsuperscript{736}Ibid, p.54.
\textsuperscript{737}Ibid: "how the actions and his identification with their springs are caused is irrelevant to the question of whether he performs the actions freely or is morally responsible for performing them" (p.54).
affirms what I earlier called the *phenomenological centrality* of identification (in opposition to Obj.A1 and A2), the passage remains agnostic on how identification is actually constituted, which implies Frankfurt’s recognition of the inadequacy of the initial model. Second, it recognizes that *however* identification is actually constituted, the states which constitute it must themselves be essentially identified with the agent, or be internal to her self without this status being conveyed to them by states of yet higher-orders (in the weak or thin sense). Third, it recognizes that this reflexive ‘identification with our identifications’ that is immediately built into our identification-conveying states is still qualitatively distinct from *decisive* identification, which is "wholehearted or complete." Fourth, while in this passage, Frankfurt uses the term "second-order volition" for these states with immediate reflexive identification that make the agent *active* in other motives, this must be understood as "second-order" in the strong or authoritative sense. Hence the term is "second-order volition" is not an *analysis* of identification: it is a placeholder for whatever state constitutes identification, since we can no longer assume that the *mere correlation* of my second-order desires with the desires on which I act establishes my identification with these desires and my responsibility for my actions.

Note that on this superseded initial model as well, whether or not such a correlation were programmed, it would still constitute identification. But this is not Frankfurt reason for holding that identification could be programmed. Rather, his argument in the passage seems to be this:

**The Quick Argument against the Origin of Identification Objection**

(1) *because* identification-conveying states (whatever they are) are themselves *inalienable* (except in the decisive sense), as the Qualitative Condition demands, and

(2) *because* identification with any motivational state constitutes responsibility for that
state, irrespective of its causal origins,

(3) we are automatically responsible for our identifications themselves (whether or not we decisively identify with them), irrespective of their causal origins.

And conclusion (3) implies that we would be just as morally responsible for our identification-bearing volitions (or volitions of the second-order in the strong sense) if they were created by manipulation and conditioning, or programmed by a D/n, because such programming would have to include our identifying with these volitions: hence "There is no paradox in the supposition that a D/n might create a morally free agent."\(^{738}\) In effect, then, Frankfurt’s argument says that the qualitative modification required to meet the regress and \(ab\) \(initio\) difficulties (Obj. B2 and B3) also shows what is wrong with the objection that the agent herself must originate her higher-order volitions, if they are really to express her inward self or identify her with motives, for action (this includes both aspects of objection B1).

Yet this ‘Quick Argument’ suffers from a question-begging suppressed premise. In order for (3) to follow from (1) and (2), we must also assume that a state’s being an identification-conveying state does not already require that it meets certain restrictions on its type of origin or history. In other words, Frankfurt assumes that the states constituting identification can be described independently of the processes through which they originate, as in Frankfurt’s initial model. Yet (as Chap. V will show) the way in which philosophers usually assure that their account of identification meets the Qualitative Condition is precisely to explain its constitution in terms of one or another set of special conditions on its origin that mere desires and preferences need not meet. The ‘Quick Argument’ against B1 manifestly begs the question against all these alternative accounts.

\(^{738}\)Ibid, p.54.
The alternative ‘historical’ approach to analyzing the constitution of identification includes limitations on origins drawn in two different ways (reflecting the two different aspects of Obj.B1), which have stronger and weaker metaphysical implications. In response only to B1.ii, an analysis may require that identification is grounded for the agent in certain sorts of considerations, dispositions, and choices between options, though without specifying the causal origin of these internal grounds. Otherwise in response both to B1.i and B1.ii, an analysis of identification may hold that identifying states are ones that originate wholly ‘within’ the agent and must therefore remain causally undetermined. More strongly still, incompatibilist historical accounts may require that identifications originate through choices that are free in the libertarian or alternate-possibilities sense. If our usual first-order motives such as desires and aversions do not have such an origin, this would be one way to establish the qualitative difference between them and identification-conveying states.

IV.4.2. From Second-Order Preferences to Strongly Evaluative Decision

Despite failing to dispense with the origin of identification issue, Frankfurt’s conception of identification clearly became more nuanced in his early papers on coercion. This development was continued in his vital though rarely cited paper on "Identification and Externality." The paper begins with the suggestion that the active/passive distinction is "considerably wider" than the distinction between actions and mere behaviors: for example, in inattentively drumming our fingers on the table, we are active without doing an action,

---

739 This is the case for Stump’s Thomistic reading of identification, as we will see in Chapter V.
740 I explore this option in the omitted additional chapter on free will.
just like a spider making unconscious but voluntary movements.\textsuperscript{742} Similarly, we can have obsessional thoughts which, though "internal" to us in one sense, are not thoughts in which we "participate actively."\textsuperscript{743} Frankfurt then suggests that, with respect to "passions" or motivational states, there is a similar active/passive distinction analogous to (though not identical with) the distinction between actions and mere behaviors: just as some movements are "strictly attributable not to the person at all but only to his body....why may a desire not, in a similar way, be an event in the history of a person’s mind without being that person’s desire."\textsuperscript{744} Although there is an obvious sense in which all of an agent’s desires, just like her body’s movements, count as "hers," this is "only a gross literal truth, which masks distinctions that are as valuable in the one case as they are in the other."\textsuperscript{745}

The identification-relation can therefore be conceived with respect to "passions" as a way of being active in them, which makes the passions "internal" to the agent or attributable to the person herself, rather than simply occurring in her mind.\textsuperscript{746} Correlatively, as the case of the unwilling addict already showed, "There is in fact a legitimate and interesting sense in which a person may experience a passion that is external to him, and that is strictly attributable neither to him nor to anyone else."\textsuperscript{747} Frankfurt considers and rejects the possibility that what makes passions "external" or alien in this way is their artificial origin, e.g. being induced by drugs.\textsuperscript{748} Then, crucially, Frankfurt considers and rejects the

\textsuperscript{742}Ibid, p.58.
\textsuperscript{743}Ibid, p.59.
\textsuperscript{744}Ibid, p.61, my italics.
\textsuperscript{745}Ibid.
\textsuperscript{746}Just as only active movements are attributed to the person herself, rather than merely to her body.
\textsuperscript{747}Frankfurt, "Identification and Externality," p.61.
\textsuperscript{748}Ibid, p.62. For as the willing addict showed, a person may provide the passion with meaning as a natural and appropriate part of his experience, irrespective of its origin: then "the person is no more a passive (continued...)
hypothesis of *his own* initial model: i.e. "that passions are external to us just when we *prefer* not to have them, or when we *prefer* not to be moved by them."\textsuperscript{749} Such preferences indicate, Frankfurt says, our ideal self-image, as when adolescents describe themselves in terms of "what they would like to be," but

the distinction between internal and external passions is not the same as the distinction between what is and what is not "real" in the sense of conforming to a person’s ideal image of himself. Surely it is possible for a person to recognize that a certain passion is unequivocally attributable to him, even when he regrets this fact, and wishes that the passion did not occur in him or move him at all. Perhaps after long struggle and disillusion with himself, a person may become resigned to being someone of whom he himself does not altogether approve.\textsuperscript{750}

This suggests the intuition, already perhaps evident in the earlier description of the unwilling addict, that the person must be *resisting or opposing* his passion in some significant way or sense for it to be external to him: otherwise, its absence may be wished, but not willed. Being preferred as part of our ideal self-image is too weak a condition for identification, because such an ideal can consist simply in *wishes* not imbued with any commitment. Thus we may not be committed in the requisite way to fighting passions within us that fail to fit this self-image.

However, there is an immediate problem with this description, for it suggests the possibility of ‘identification through resignation,’ as we might term it. Yet in his later paper on action, Frankfurt claims that the case of the unwilling addict shows that

the attitudinal conditions of a person’s action may themselves be alien to him. There is no reason to assume that an addict who succumbs unwillingly to his cravings finally adopts as his own the desire he has tried to resist. He may in the end merely submit to it with resignation, like a man who knows he is beaten and

\textsuperscript{748}(...continued)
bystander with respect to it than if it had arisen in more integral response to his perceptions" (p.62).
\textsuperscript{749}Ibid, p.63, my italics.
\textsuperscript{750}Ibid, p.64.
who therefore despairingly accepts the consequences defeat must bring him, rather than like someone who decides to join with or to incorporate forces which he had formerly opposed.\textsuperscript{751}

There is, then, a kind of "acceptance" of our passion which is entirely consistent with continuing to oppose it in the sense necessary for it to remain "external" to us. Let us call this ‘acceptance in protest’ or ‘protesting resignation.’\textsuperscript{752} This contrasts sharply, for example, with a case in which an unwilling but unreformed addict takes her failure to beat the addiction as grounds for openly embracing it, perhaps in order to avoid the pain and difficulty of battling it, and thus becomes a willing addict. In such cases, a person may change her identification itself to conform with her dominant first-order desires, even though they still conflict with her ideal self-image. But in that case, she is ‘resigned to them’ not in protest but in active self-commitment to them; she does not merely acquiesce out of necessity, but invests herself in these desires, and thus to the extent that she still regrets or does not approve them, her "regret" is reduced to the status of a wish: it lacks the force of will.

The qualitative difference between identification and preferences (or desires in the S2 sense) must therefore depend on positive self-commitment, and not merely in protesting resignation or acquiescence. With this proviso, we can appreciate the development in Frankfurt’s analysis: since attitudes such as preferences for passions can simply express wishes related to an ideal self-image, "whether a passion is internal or external to a person is

\textsuperscript{751}Frankfurt, "The Problem of Action," p.77.

\textsuperscript{752}I hesitate to use Frankfurt’s term —"despair"— for this kind of acceptance, because existentialist authors (particularly Kierkegaard in the Sickness Unto Death) usually use this term for a variety of states that involve either active identification with certain passions or else active alienation of them. The unwilling addict is not necessarily in "despair" in any of the existential senses which Kierkegaard distinguishes, however. Rather, this addict "accepts" her addictive desires only in the sense of recognizing that, willy nilly, they will move her to action, and trying to salvage what she can from this less than optimal situation.
not just a matter of the person’s attitude toward the passion.”\textsuperscript{753} His attitudes towards it may still be relevant, since contented acceptance of a frequent emotion or desire may indicate that the person "has done whatever it takes to identify himself with it." Disapproval of a passion may be a necessary condition for its being external to the person, but it is not by itself a sufficient condition.\textsuperscript{754} Identification and alienation (or "internality" and "externality") cannot "be explicated simply in terms of a person’s attitudes," as the initial account in terms of second-iteration desires has proposed, because such attitudes themselves are susceptible to externality:

Suppose that a person has mixed feelings about one of his passions: he is aware of having an inclination to approve of the passion and also of having an inclination to disapprove of it. Suppose he resolves this conflict by decisively adopting an attitude of disapproval toward the passion. He may find nonetheless that his inclination to approve the passion persists, though it is now external to him...

The fact that a person has a certain attitude toward a passion can be construed as determining either the internality or the externality of the passion, surely, only if the attitude in question is itself genuinely attributable to him. An attitude in virtue of which a passion is internal, or in virtue of which a passion is external, cannot merely be an attitude that a person finds within himself; it must be one with which he is to be identified.\textsuperscript{755}

An attitude such as a desire of the ‘second-order’ (in the weak sense) may persist without constituting an identification-bearing and conveying volition. This again illustrates Frankfurt’s recognition of the Qualitative Condition: identification and alienation can be constituted only by states that are themselves inalienable, or essentially internal to the agent: since preferences and evaluative attitudes about desires are not inalienable, "an infinite regress" arises if we try to explain identification in terms of them alone.\textsuperscript{756} And as a result,

\textsuperscript{753}Frankfurt, "Identification and Externality," p.64.
\textsuperscript{754}Ibid, p.65.
\textsuperscript{755}Ibid, pp. 65-66, my italics.
\textsuperscript{756}Ibid, p.66.
there is a clear shift towards affirming that identification is something we do, e.g. either as a second-order action or as a disposition to second-order acts. It is this active or ‘volitional’ quality\(^757\) which makes identification qualitatively distinct from all alienable attitudes, such as inclinations or preferences for other desires.

In addition, Frankfurt also suggests a new partial analysis of the kind of act that constitutes identification, which further distances it from the latent emotivism of his initial view that identification may lack any objective evaluative content and be "thoughtless and spontaneous" in controlling the first-order will.\(^758\) This analysis depends on a noteworthy phenomenological difference between two kinds of "conflict of desire." In the first, the "desires at issue" are like mere preferences, e.g. to see a film or watch a concert. Their conflict can therefore be resolved by arranging them into a single ordering, like preferences in standard rational decision theory, which means that if seeing the film is preferred to seeing the concert, but it turns out that the agent cannot get into the film, "Then it would be quite natural for him to revert to his second choice and go to the concert."\(^759\) In preferring the first to the second option, the agent does not cut himself off from the second option, but retains it as a fallback.

But in the other kind of conflict, resolution requires that the agent "reject one of the desires that he feels altogether," or place it entirely outside the ordering containing his other option.\(^760\) Such resolution thus decides between completely incommensurable alternatives. As a result: for example, in deciding between a generous desire to compliment an

\(^{757}\) In Pink’s sense of ‘will’ as second-order agency, then, identification is an act of will. However, I have yet to show that identification is volitional in my more specialized sense: namely, that it is constituted by second-order acts or dispositions that are projectively motivated.


\(^{759}\) Frankfurt, "Identification and Externality," p.66.

\(^{760}\) Ibid, p.67.
acquaintance and a "spiteful desire to injure the man," there can be no fallback: if he does not get the chance to compliment him, "This would not naturally lead him to see if he can salvage the satisfaction of his other desire."\(^{761}\) Since these desires are not options within a single preference-ordering, they will arise either such that one is already identified or internal and the other alienated, or such that both are ambiguously identified: then "the person himself is in conflict and not merely the desires that are occurring in his mental history."\(^{762}\) In neither case does either desire appear as simply neutral, or as a passion with respect to which the agent is wanton. This example suggests that it is characteristic of motives with which we identify —and thus of identifications themselves— that they can conflict with others (also either identified or alienated) only in a strong fashion which requires them either to be accepted or rejected outright, rather than simply assigning them a weight greater or less than the weight of other options. Motives we can weigh this way are mere preferences. This implies (though Frankfurt does not put it this way) that identification cannot consist in simple weighing of first-order motives. This accounts for at least part of its qualitative difference from S2 desires.

It is interesting that there are now technical considerations in the theory of preferences in rational decision and game theory which point in the same direction. For example, Steedman and Krause write that because of their capacity for "self-evaluation involving second-order desires," it turns out that "humans are too complex for their actions to be

\(^{761}\)Ibid, p.67. One can easily see that Frankfurt’s unwilling addict is an illustration of an agent in such a predicament. He rejects identification with his addictive desires. But on finding no success in resisting these desires (i.e. on discovering that they continue to be the will on which he acts, in spite of his inward self), this addict does not thereby convert to a willing addict (if he does, it is a separate movement). His attitude is not one of merely preferring non-addictive desires to addictive desires, such that if the latter turns out to be impossible for him, he will willingly go with the latter as a second-best option.

\(^{762}\)Ibid, p.67.
predictable from complete, transitive preference-orderings. They remarks imply that second-order "preferences" cannot be part of a single complete linear ordering of preferences along with the first-order preferences over which they dispose, since a "single-preference ordering representation" including first- and second-order preferences leads to paradoxes. If this is right, then the very idea of second-order preferences — in the same minimal S2 sense of ‘preference’ that rational decision theorists use in defining ordinary first-order preferences — is not even coherent: it leads to diagonalization paradoxes. In that case, our second-order attitudes towards first-order preferences and desires cannot just be more preferences, at least in the minimal sense of preference defined within the formalisms of rational decision theory. Rather, these attitudes must be something qualitatively distinct from simple weights or rankings of first-order motivational states.

Instead of second-order preferences, then, identification involves what Charles Taylor has called "strong evaluation" of the desires on which we might act (see Chap. II, §3.8). For it is not hard to see that the two types of conflict Frankfurt contrasts in "Identification and

---


764 Ibid, pp.225-226. They also cite Nozick’s argument in his Philosophical Explanations that decisions do not simply follow from pre-weighted considerations and options, but assign the weights to competing reasons: "To the (non-negligible) extent that Nozick is right, the ‘Arrow problem’ for the individual, Levi’s insistence on decision without full pre-reconciliation, the infinite regress problems about choices over choices and decisions subject to information costs, the constant complaints that Pareto could not just replace agents by their preference maps — all fall into a consistent pattern. Decision is not fully representable ex-ante as constrained optimization over a complete, transitive or even pre-given ordering" (p.226).

765 There is something like a ‘Third-Man’ dilemma for preferences here. Suppose that P,a and P,b are first-order preference, which rank options in different ways: P,a ranks option A over B, and P,b ranks B over A. What then could it mean to have a second-order preference P₂ for P,a over P,b, except that one ranks A over B, in which case P₂ is identical to P,a. If this is all that is needed to generate a higher-order preference, then having any preference generates an endless series of preferences for this preference, etc. But if instead a second-order preference must itself be a discrete element, distinct from the preferences, it is about, then it must somehow be a ranking of rankings, which is not simply a ranking of options for action. But then our preferences do not form a single scale or ranking, as standard rational choice theory requires. For what could it mean to rank A-over-B over ranking B-over-A other than to rank A-over-B again?
Taylor’s idea that strong evaluation considers options which are essentially rather than contingently contrasted helps explain why such options cannot be arranged in a single preference-ordering: this is why identification with one option rules out identification with its contrary not contingently, but categorically. At the same time, Frankfurt’s contrast between choices where the rejected option is a fallback and those in which it is not help demonstrate the significance of Taylor’s contrast, underlining the real difference between weak and strong evaluations. Yet in this essay, Frankfurt does not claim that identification is constituted simply by strong evaluations either, for it also involves decision:

..the person renders the second desire external to himself and identifies with the first. Here it appears to be by making a particular kind of decision that the relation of any person to his passions is established. It may be that a decision of this kind, even when it is not visible as in the present example, lies behind every instance of the establishment of the internality and externality of passions.\textsuperscript{767}

The reason for appealing to decisions about motives, as possible constitutive elements of identification becomes clear in a footnote, where Frankfurt remarks that ”Decisions, unlike desires or attitudes, do not seem to be susceptible both to internality and to externality.”\textsuperscript{768} In other words, decisions seem to meet the Qualitative criterion, because they are inalienable (at least at the time when they are made).

**IV.4.3. Care and Decision**

In subsequent papers, Frankfurt reformulates this notion of evaluative "decision" involved
in identification, in part because of fears that its voluntarist connotations could lead to familiar problems of arbitrariness. Thus in "The Importance of What We Care About," Frankfurt turns to the notion of caring about something as guiding oneself by reference to it. In Frankfurt’s sense, caring is "essentially a reflexive" activity, in which the person "identifies himself with what he cares about." As I described earlier (Chap. III §2), caring involves a kind of "investment" of oneself in ongoing pursuits that would be meaningless if it were only momentary. By itself, however, this might suggest that care consists in nothing more than persisting patterns or dispositions of first-order motivation. After all, we do not usually describe our cares as being explicitly about our own motivational states. But Frankfurt conceives caring as the process by which a person "fashion[s]" the "themes and structures" of his life as a whole, including his passions: since developing the right sort of affections and desires is intrinsic to caring about another person, for example, attitudes towards our own motivational states are involved in caring about them. In the process of caring about family, political causes, hobbies, institutions, sports, etc., a person gives the sequence of her first-order passions a first-personal significance beyond her merely having them: she "construes them as being bound together" by fitting her cares, or alternatively, as not fitting in because they conflict with what she really cares about.

---

769Frankfurt, “The Importance of What We Care About,” *Synthese*, 53.2 (1982); reprinted in *The Importance of What We Care About*: 80-94, p.82.

770Ibid, p.83.

771Conflict among cares is existential (rather than wanton) precisely because it is a conflict of investments of this sort: the person himself is committed in an enduring sense to incompatible things.

772Some of Frankfurt’s descriptions make it sound as if care is simply a state of first-order affectivity: caring for family traditions or mathematical truth, for example, sound like first-order desires (ibid, p.81).

773Ibid, p.83.

774Ibid, p.83.

775Thus if we think of the raw pattern of actual emotions and desires, a person experiences, the pattern which (continued...)
As a result, we cannot think of caring about things or persons merely as a composition of more primitive states, or as "a complex set of cognitive, affective, and volitional dispositions and states," as Frankfurt once suggested.\textsuperscript{776} Since what is distinctive of care is not simply having a set of first-order states over time, but regarding them as important because they fit with concern for the object of one’s care, caring clearly involves reflexive or weakly-higher order attitudes, though it is not simply constituted by higher-order desires or preferences. Instead (as I briefly suggested in Chapter III) care in Frankfurt’s sense must involve identification with the other states it requires: the volitional component is what unifies the whole complex of states involved in completely realized caring.\textsuperscript{777}

In two places, Frankfurt recognizes this close association between identification and "caring" in his sense. In his first essay on care (discussed in Chapter III), Frankfurt says that there is a sense in which a person "identifies himself with what he cares about."\textsuperscript{778} In a more recent essay, Frankfurt repeats this formulation, after noting that "Caring is essentially volitional" (as opposed to its cognitive and affective grounds).\textsuperscript{779} He adds that, unlike mere desires and impulses, "The personal characteristics of someone’s will are reflexive, or

\textsuperscript{775}(...)continued

constitutes his cares will be made up of elements selected out of the raw first-order pattern. We could then say that the agent identifies with those included in the inner pattern, while those in the raw pattern but excluded from the inner pattern are alienated.

\textsuperscript{776}Ibid, p.85. And indeed Frankfurt alters this formulation in later papers such as "Autonomy, Necessity, and Love."

\textsuperscript{777}And as a result, more attenuated forms of caring are possible in which one identifies with the right combination of first-order states but has not yet actually developed them all. It may be helpful to think of such a person as trying to care about something: his ‘heart’ is in it, though the rest of him is not yet there. This differs of course from cases where a person desires to develop certain interests, concerns, and involvements (perhaps out of desire to please someone else) but her heart is not in it, or she does not identify with these states even if (like the person who cares in the complete or ideal sense) she succeeds in developing some of them. Such a person has a semblance of care. She may in addition deceive herself into thinking that she cares in the way these interests and concerns normally suggest, but an effort is required to maintain such an illusory self-image, and such an effort can always be thwarted by events that will disillusion the agent.

\textsuperscript{778}Frankfurt, "The importance of what we care about," p.83.

higher-order, volitional features. They pertain to a person’s efforts to negotiate his own way among the various impulses and desires by which he is moved.\textsuperscript{780} This useful description suggests both how caring involves identification with some of our (potential or actual) first-order states and alienation of others, and that identification is an intrasubjective or (weakly) second-order stance towards one’s own states, even if it is not constituted by its hierarchical position.

As Stump has recently noted, however, the description of "caring" does not provide a reductive analysis of identification or being active with respect to our motives, in the way that Frankfurt’s initial model attempted, because the definition of caring involves reference to the notion of identification or being active.\textsuperscript{781} The care model provides instead a suggestive way of describing the phenomena of identification.\textsuperscript{782}

This description of identification in terms of care is supposed to do three things for Frankfurt. First, because caring is distinguished qualitatively from first-order desires and beliefs, which "have no inherent persistence"\textsuperscript{783} and only require a succession of moments, as in Hume’s portrait of the ‘self’ as a bundle, care displays the dynamic features of duration and continuity, which seem required to resolve the objection that second-order volitions could be sporadic (Obj.B4). Identification with one’s first-order states cannot occur intermittently through a coruscation of brief and unconnected second-order desires; rather, identification requires cares that are themselves enduring and give one an inward character, tending to the establishment and preservation of stable first-order complexes (or outward

\textsuperscript{780}Ibid, p.23.


\textsuperscript{782}Nevertheless, when care itself is understood in terms of projective motivation, it may provide a constitutive analysis of volitional identification, as I will argue below.

\textsuperscript{783}Ibid, p.84.
character) at the same time. This in turn means, as suggested in introducing Obj.B4, that we
identify not with highly particular motives, (or those individuated by very short term
temporal segments) but rather with larger features of an entire ‘outward character’ (in this
sense defined in IV §I). As Frankfurt says subsequently,

...It is a salient characteristic of human beings, one which affects our lives in deep
and innumerable ways, that we care about what we are. This is closely connected
both as cause and as effect to our enormous preoccupation with what other people
think of us. We are ceaselessly alert to the danger that there may be discrepancies
between what we wish to be...and how we actually appear to others and to
ourselves.

We are particularly concerned with our own motives. It matters greatly whether
the desires by which we are moved to act as we do motivate us because we want
them to be effective in moving us or whether they move us regardless of ourselves or
even in spite of ourselves.784

This clarifies both that care involves reflexive (or ‘weakly’ higher-order) attitudes towards
our motives, and that the motives these attitudes pick out are understood as parts of the
stable outward character we care about expressing in our actions over time.785

Second, as a result, Frankfurt can argue that "the fact that a person decides to care about
something cannot be tantamount to his caring about it."786 Instead:

If we consider that a person’s will is that by which he moves himself, then what he
cares about is far more germane to the character of his will than the decisions or
choices he makes. The latter may pertain to what he intends to be his will, but not

784Frankfurt, "Identification and Wholeheartedness," p.163, my italics. The "because" indicates that we want
our identifications to be effective in guiding our first-order motives, or to have free will in Frankfurt’s sense.

785This does not mean that it would be impossible (though it would be unusual) for someone to identify with
a certain variation in motives or dispositions that would confuse third-party interpretations of her character.
Nor do I mean to imply that a person with cares cannot be totally oblivious to what others think of him (or even
care about not taking how they will view him into account). Nevertheless, the features with which he identifies
himself are of the type that make up an outward character on which others most naturally and immediately
focus in interpreting the agent’s actions over time: e.g. they are features like being prone to anger, rather than
being angry at a particular insult; or needing praise and encouragement, rather than desiring the boss’s approval
for performance on a particular assignment today; or being sympathetic to suffering, rather than desiring to help
alleviate the pain of victims of a tragedy in a particular town on a given date, etc.

786Frankfurt, "The importance of what we care about," p.84.
necessarily to what he truly is.\textsuperscript{787}

In other words, the phenomena of care, which include "several varieties of love,"\textsuperscript{788} represent the agent’s essential activity, or exhibit the immediate and necessary identification with the agent required by the Qualitative Condition, better even than his decisions do.

This is a third reason for introducing the concept of care: the "activity of caring as such," taken apart from the importance of the objects of care (which is partially given to them by our caring about them), has value for us just because it is through caring that we actively shape our character, or determine what kind of person we are.\textsuperscript{789} Caring about something is \textit{inalienable} in a way that unifies the self over time, while simply deciding to care may fail, because the agent remains "\textit{unable} to carry out his intention" and does not develop the "feelings, attitudes, and interests constitutive of the sort of person which his decision has committed him to being."\textsuperscript{790} Caring is an \textit{actual} accomplishment.

However, the persuasive point here is that caring is an entrenched volitional state that by its nature cannot normally be altered just ‘at will:’ even apparently revolutionary changes in our cares must be foreshadowed by internal developments, whether we are aware of them or not. But this does not imply that decisions have \textit{no role} either in maintaining or in eventually changing our cares. Once we see that care functions in effect as a new analysis of \textit{identification} for Frankfurt,\textsuperscript{791} and one meant to address recognized deficiencies with his

\textsuperscript{787}Ibid, p.84.
\textsuperscript{788}Ibid, p.85.
\textsuperscript{789}Ibid, p.93.
\textsuperscript{790}Ibid, p.85.
\textsuperscript{791}Though, as I said, it is not an analysis in the reductive sense, and though Frankfurt seems intentionally not to stress the link between care and identification in "The Importance of What We Care About" —as if he had come to doubt the usefulness of the concept of "identification" at this point and sought to \textit{replace} it with the concept of care (though he returns to identification in later writings).
initial model, his account of it seems to dismiss the role of decision a bit too quickly and to leave unclear the relation of the care analysis to the evaluative decisions described in "Identification and Externality." Frankfurt acknowledges this link in a more recent statement:

To be a person entails having evaluative attitudes (not necessarily based on moral considerations) toward oneself. A person is a creature prepared to endorse or repudiate the motives from which he acts and to organize the preferences and priorities by which his choices are ordered. He is disposed to consider whether what attracts him is actually important to him. Instead of responding unreflectively to whatever he happens to feel most strongly, he undertakes to guide his conduct in accordance with what he really cares about.  

Here Frankfurt differentiates caring from desiring in part by appeal to the point made in "Identification and Externality:" our identification-bearing states decide between incommensurable preference-hierarchies, and thus personal agents do not merely act in accord with pre-given preference-orderings.

In this light, it seems that the readiness to make such decisions must be part of what it means to care about something. For example, caring about my wife surely requires that I be disposed to help her with a Saturday event that is important for her career, if I can. And this means that I am ready to rule out conflicting possibilities such as disclaiming any concern for her career, because I strongly evaluate the former as the right option. So strongly evaluative decisions or dispositions to them (covering a range of possible counterfactual circumstances) are involved in caring. The account in "The importance of what we care

793 Thus if it turns out that her event is canceled, I might (say) spend my Saturday watching football instead, but this would not be because I just preferred her event to football in a simple weighing and fell back on my second-best option when her event was canceled. Rather, my commitment to helping her be successful remains in the background, ready to trump conflicting alternatives if another way of fulfilling the same end comes up, e.g. her boss invites us out to dinner in lieu of the canceled event. What conflicts with my commitment essentially is the alternative of not giving a darn about her career, and I cannot simply fall back on this alternative if there is no way to aid her, because I do not simply prefer the former alternative to the latter.
about" can thus be linked to the account given in "Identification and Externality" if we think of cares as strongly evaluative dispositions: e.g. caring about animals would involve a settled propensity (a) to evaluate certain motives, such as the desire to preserve natural habitats unspoiled) as appropriate or indispensable within a wide range of circumstances, and (b) to decide on this basis to rule out contrary motives in the strong sense of regarding them as 'alien’ or never a good second option. On this analysis, cares are (or at least involve) dispositional or dynamic versions of the evaluative decisions described in "Identification and Externality."

Thus Frankfurt’s contrast between care and decision should not be understood as meaning that the two are simply unrelated: rather, the point only seems to be that a sort of higher-order decision, namely the decision to care, is not by itself sufficient for caring (or sufficient to guarantee that the agent develops the required disposition to appropriate evaluative decisions about first-order intentions). It is an additional and distinct point that neither the decisions already involved in caring, nor the decision so to care, will guarantee that the agent 'decisively’ or completely identifies with his cares (as we will see below, Frankfurt tends to run these points together).

In that case, although we are active in our cares, caring would not itself be an "act" in the narrower sense of something that is the object of intentions formed by decisions: "The formation of a person’s will is most fundamentally a matter of his coming to care about certain things, and of his coming to care about some of them more than about others." But "coming to care" is not just deciding to care. It may be right that this (third-order) activity of forming cares cannot be understood as an analog of the decisions by which form our

---

794 Ibid, p.91.
everyday intentions, to act, as (for example) Pink has described them. Yet even at this level, it is essential to Frankfurt’s picture that caring (or not caring) about something cannot simply happen to us, as first-order desires and addictions may occur. If the person’s "decision" to care about X fails, as Frankfurt says that it may, her "failure" to develop the volitional dispositions, feelings, attitudes, and interests involved in caring about X cannot be like the unwilling addict’s discovery that despite his volitional identification with healthy desires, the desire, to shun drugs does not develop in accordance with his will. Rather her decision to care about X can fail not because it encounters an external obstacle (either inside or outside her mind), but only because it conflicts with other states internal to her will, i.e. other cares which she cares about more. In other words, if a decision to care fails, it must be an essential rather than contingent failure.

Frankfurt’s own analysis tends in this direction when he re-affirms the role of decision in "Identification and Wholeheartedness." In this essay, he again describes decision as ordering preferences within a hierarchy and excluding some absolutely through strong evaluation: "It is these acts of ordering and of rejection—integration and separation—that create a self out of the raw materials of inner life." Moreover, as the care analysis implied, the identification which involves (and in part occurs through) such inward decisions forms an inner character by appropriating or taking up features of the agent’s "habitual dispositions"

---

795 Thus it was a bit misleading for Frankfurt to say that "A decision to care no more entails caring than a decision to give up smoking entails giving it up" (p.84), for this suggests the very parallel he needs to deny.

796 The disposition of this question thus ultimately turns on whether or not persons can come to care about something necessarily, or because it is simply essential to their identity, without anything like a voluntary directing of oneself one way rather than another in the process. This is to raise, with respect to cares, the very same question of genesis raised with respect to identifications by Obj.B1.i. If this is impossible, as I will argue in the projected additional chapter on free will, then an existentialist can say that when decisions to care fail, it is because they encounter contrary commitments in which the agent has made something like a choice between alternative commitments.

of first-order motivation. Thus, contrary to Aristotle’s analysis [in *Nicomachean Ethics*, III.5]:

The fundamental responsibility of an agent with respect to his own character is not a matter of whether it is as the effect of his own actions that the agent has certain dispositions to feel and behave in various ways. That bears only on the question of whether the person is responsible for having these characteristics. The question of whether the person is responsible for his own character has to do with whether he has taken responsibility for his characteristics. It concerns whether the dispositions at issue, regardless of whether their existence is due to the person’s own initiative and causal agency or not, are characteristics with which he identifies and which he thus by his own will incorporates into himself as constitutive of what he is.798

Note the similarity of this characteristics/character distinction to my own earlier distinction between inner and outer character.799 Though we might suspect that identifying with characteristics would tend to deepen their entrenchment or help strengthen their influence on the motives,₁ of our actions, identification does not consist in this kind of causal effect since, as Frankfurt insists, the alienated motives,₁ may persist.800 In that respect, there may still be conflict, but identification transforms it from a conflict "between the desires,₁" into "a conflict between one of them and the person who has identified himself with its rival."801

To accommodate his earlier concerns, Frankfurt then distinguishes deciding from choosing: decision is reflexive agency whereas the object of choice is something other than

798Ibid, pp.171-172. Note that the phrasing here implies that being responsible for having characteristics is something different from taking responsibility for these characteristics themselves. Although this remains somewhat unclear, it seems that in the former case, the agent would simply be responsible for the characteristics as effects of his action. But being responsible for them as effects is not enough for them to have the first-personal meaning of a coherent outward character for him: this requires taking responsibility for them by identifying with them.

799Frankfurt’s distinction is not exactly the same as mine, however, since in his terms, our "characteristics" (outward character) become part of our "character" when we internalize or incorporate them. In my terms, our "inward character" consists in our identification with some of our (actual or potential) Frankfurtian characteristics, and thus has as its content what Frankfurt calls "character." Inward character is thus a persisting intrasubjective or reflexive volitional state of the agent.

800Ibid, p.172.

801Ibid, p.172.
our agency:

This difference between deciding and choosing accounts for the fact that deciding to make a certain choice is not the same thing as actually making it...whereas deciding to make a particular decision cannot be distinguished from making the decision itself.802

Decisions are in this sense "iteratively self-implementing," which is connected to their being actions, like raising my arm (as opposed to its rising).803 But though a person may decide to resolve his ambiguity by identifying wholly or completely with one type of motive, thus trying "to supersede a condition of inner division and to make himself into an integrated whole," the attempt may fail because she still has cares that involve other motives.804 In that case, Frankfurt says, this must be because the agent still "has other intentions," whether recognized or not, and thus her decision to identify purely with the first type of motive was itself "not wholehearted," or not one with which the agent completely identifies.805

If this is right, it suggests that, while our decisions are inalienable in the sense that we are identified per se with them, and the evaluative decision to identify with a motive may be essential to positive identification with it (as proposed in "Identification and Externality"), such decisions do not automatically carry complete or "decisive" identification with them,
since they may coexist with contrary decisions. Thus the disposition to a given kind of strongly evaluative decision about motives, may (at least partially) constitute caring about or positive identification with certain motives, but such dispositions will not by themselves be sufficient to constitute wholehearted identification, since persons are even capable of having conflicting strongly evaluative dispositions or cares. In a sense, then, "decisive" was an inappropriate term to describe complete or unambiguous identification, which in part explains Frankfurt’s shift to the language of "wholehearted" identification.

In sum, these developments in Frankfurt’s analysis since his "Freedom of the Will" paper have gone quite a way towards addressing the main objections to his initial model by describing identification as: (1) temporally extended in its process, (2) having enduring character rather than simply to the fine-grained motives of highly specific acts as its object, and (3) as qualitatively distinguished in its intensional attitude from mere desires or serial preference-orderings by the strongly evaluative content of decisions the disposition to which is involved in identification. Thus decisions based on a strong evaluation of motives, help integrate a person
dynamically insofar as it provides...for coherence and unity of purpose over time; statically insofar as it establishes —in the way discussed earlier—a reflexive or hierarchical structure by which the person’s identity may be in part constituted.806

Despite this real progress, however, certain problems remain for Frankfurt’s revised account, both with respect to Obj. B1 and with respect to the issue of decisive identification.

IV.5. A First Approach to Decisive or ‘Wholehearted’ Identification

In "Identification and Wholeheartedness," Frankfurt substantially recasts his initial

806Ibid, p.175.
account of decisive identification in response to Obj.B5. But he also now portrays this explanation of decisive identification as a way to "deal with" th[e] problem about identification per se, which he recognized in this essay: namely, that positive identification cannot be constituted simply by second-order desires with a certain type of content.\textsuperscript{807}

Accordingly, there are two issues for discussion in this section. The first (more general) issue concerns the proper function of an account of decisive identification in the phenomenology of identification as a whole. I will argue, in contrast to Watson and Frankfurt, (a) that positive identification and decisive identification concern different phenomena; (b) that the former must be psychologically more fundamental; (c) that the function of an account of decisive identification should be to solve Obj. B5, which is a separate problem from the regress and ab initio problems facing identification per se; (d) and that the mistaken belief that an account of decisive identification can resolve Obj. B2 and B3 by providing the basis for a new account of the constitution of identification per se derives from a conflation of the problems and a misconstrual of what is needed to resolve the regress paradox.

The second (more specific) issue concerns whether Frankfurt’s new account of decisive identification is convincing in its own right, i.e. as a way of meeting Obj.B5 rather than as a basis for a qualitative account of positive identification itself. With respect to this question, I will also argue that Frankfurt’s new account of decisive identification or "wholeheartedness" in terms of "satisfaction" remains inadequate in important ways.

\textbf{IV.5.1. The Proper Function of an Account of Decisive Identification}

\textsuperscript{807}Frankfurt, "Identification and Wholeheartedness," p.166. For the description of his recognition of Obj. B2-B3, see the end of §3.2 above.
A certain confusion about the role of decisive identification within a theory of identification-phenomena in general enters the history of this subject with Gary Watson’s formulation of the Arbitrariness objection (B5). Watson presents the objection in two remarks. First, he says that in cases of conflict,

Since second-order volitions are themselves simply desires, to add them to the contest of conflict is just to increase the number of contenders: it is not to give a special place to any of those in contention. The agent may not care which of the second-order desires win out. The same possibility arises at each higher order.

Watson regards this point as seamlessly linked with the basic problem that being about other desires does not make second-order volitions authoritative. Yet this difficulty, as formalized in Obj. B2-B3, is a problem with explaining how identification is constituted so as to position the self relative to conflicts among first-order desires. But Watson then implies that the conflict he has in mind in the above passage is between second-order desires. He confirms this by turning next to Frankfurt’s suggestion that decisive identification resolves such higher-order conflicts by "resounding" through all potential higher orders of volitions. In reply to Frankfurt, he remarks:

We wanted to know what prevents wantonness with regard to one’s higher-order volitions. What gives these volitions any special relation to ‘oneself?’ It is

---

808In fairness, this confusion is also evident in Peggy Nicholson’s article, “Decisive Identification and the Concept of a Person,” which predated Watson’s by a few months. As Nicholson says, in “Freedom of the Will and the Concept of a Person,” Frankfurt “does not clearly state whether identification through the formation of a second-order volition is always decisive or not” (note 1, page 293), but she treats the problem for decisive identification as identical to the problem for identification per se. Thus she gives a version of the ab initio problem that applies to decisive identification: if an individual’s "decisiveness" consists in his simply having a second-order volition that he never questions, which thereby "puts an end to questions about higher-order desires," then it is impossible to see why a wanton is worse off. The problem is the weak sense of "decisiveness" employed here: it means simply having an n-th order will which closes off consideration of n+1 order volitions which could potentially alienate one’s n-th order volition. In this recursive sense, the wanton can also be said to have "decided" not to consider any desires of a higher-order than the one which, by default, forms his will. The wanton is simply being decisive, if decisive identification only requires having a will of some order which "resounds" through all higher orders, in the weak sense of closing one’s eyes to alternatives at higher-order levels. She also poses a version of Obj.B5 against this sense of decisiveness: "Questions about higher order desires may be quite significant, even though the person never allows himself to consider them" (p.292) —which makes the decisive agent’s pre-emptive refusal look fanatical.
unhelpful to answer that one makes a ‘decisive commitment,’ where this just means that an interminable ascent to higher orders is not going to be permitted. This is arbitrary.\textsuperscript{809}

This presupposes, I think incorrectly, that the proper function of the initial account of decisive identification was to explain how higher-order volitions become \textit{to any extent} authoritative, or constitute identifications \textit{at all}. But in fact, its function was initially to answer a \textit{different} though deceivingly similar ‘regress problem’: namely, how \textit{potential} conflicts among second-order volitions can themselves be resolved without simply postulating third-order volitions that could also be in conflict, and so on. Frankfurt first introduced the issue of decisive identification in the context of recognizing that "people are generally far more complicated" than his sketch of first- and second-order desires suggested.\textsuperscript{810} In his example, Frankfurt suggests that a person may decisively identify with his desire\textsubscript{1} to concentrate on his work, because his volition\textsubscript{2} to act on that desire\textsubscript{1} is accompanied by the decision "that no further question about his second-order volition, at any higher-order, remains to be asked."\textsuperscript{811} This obviously could not be the case if the agent also had a volition\textsubscript{2} opposed to this desire, and so this decision would have to \textit{rule out} such opposing volitions\textsubscript{2}.

But ever since Watson portrayed this as an arbitrary termination of "\textit{the} regress problem," the different issues at stake in Objections B2 and B5 have been muddled together. This confusion gets started because, as we saw above (§2), Frankfurt himself is not initially clear about the \textit{asymmetry} of the volition/desire distinction between orders: i.e. that a second-order volition does not mean our single \textit{effective or strongest} second-order desires. Hence


\textsuperscript{811}Ibid, p.22.
And of course, if second-order volitions were formed from second-order desires in a fashion precisely analogous to the way that first-order volitions emerge from among one’s first-order desires just by becoming the strongest, then naturally it would look like the person could just sit back and wantonly wait to see which second-order desire constitutes her identification by winning out, without caring which it is. Moreover, if she did care, it would then be by way of a third-order volition. But if this is simply whichever third-order desire is strongest, the problem repeats itself \textit{ad infinitum}. What reformulating the problem this way shows is again that to be identification-bearing, higher-order volitions must be qualitatively distinct from first-order volitions, and in this light, we can see that the asymmetry in Frankfurt’s initial model (which refuses simply to equate second-order volitions with whichever second-order desire is strongest) is really a nascent anticipation of this need for qualitative differentiation, or a first step in the right direction.

Moreover, it is worth noting, this is why it is much more important than it might appear to refuse Eleonore Stump’s suggestion that we should make a neat symmetric parallel by revising Frankfurt’s model so that a second-order volition is just an “effective second-order desire...which moves the agent all the way to the action of making the corresponding first-order desire his will.” This has the result, as she notes, that the reformed unwilling addict has a second-order volition not to act on the desire for heroin, but "the unwilling addict who succumbs to his addiction” only has this as a second-order desire (see Stump, "Sanctification, Hardening of the Heart, and Frankfurt’s Concept of Free Will," p.217). This introduces several confusions: aside from apparently denying the unsuccessful unwilling addict any \textit{identification} with his healthy desire, it also conflates identifying with a desire and having \textit{free will} in acting on whatever desire is our will,—states which Frankfurt meant to distinguish, as we saw.

Watson is able to misinterpret Frankfurt as saying that identification is constituted by one of the agent’s second-order desires ‘winning out;’ thus Watson equates an unresolved second-order conflict with a state of wantonness about one’s higher-order volition. \footnote{And of course, if second-order volitions \textit{were} formed from second-order desires in a fashion precisely analogous to the way that first-order volitions emerge from among one’s first-order desires just by becoming the strongest, then naturally it would look like the person could just sit back and wantonly wait to see which second-order desire constitutes her identification by winning out, without caring which it is. Moreover, if she did care, it would then be by way of a third-order volition. But if this is simply whichever third-order desire is strongest, the problem repeats itself \textit{ad infinitum}. What reformulating the problem this way shows is again that to be identification-bearing, higher-order volitions must be qualitatively distinct from first-order volitions, and in this light, we can see that the asymmetry in Frankfurt’s initial model (which refuses simply to equate second-order volitions with whichever second-order desire is strongest) is really a nascent anticipation of this need for qualitative differentiation, or a first step in the right direction. Moreover, it is worth noting, this is why it is much more important than it might appear to refuse Eleonore Stump’s suggestion that we should make a neat symmetric parallel by revising Frankfurt’s model so that a second-order volition is just an "effective second-order desire...which moves the agent all the way to the action of making the corresponding first-order desire his will.” This has the result, as she notes, that the reformed unwilling addict has a second-order volition not to act on the desire for heroin, but "the unwilling addict who succumbs to his addiction” only has this as a second-order desire (see Stump, "Sanctification, Hardening of the Heart, and Frankfurt’s Concept of Free Will," p.217). This introduces several confusions: aside from apparently denying the unsuccessful unwilling addict any \textit{identification} with his healthy desire, it also conflates identifying with a desire and having \textit{free will} in acting on whatever desire is our will,—states which Frankfurt meant to distinguish, as we saw.}

\footnote{Ibid, p.108.}
first way, since her identifications are only supposed to consist of second-order desires (with a content that makes them second-order volitions). If this were right, then resolving Obj. B2 and B3 should clear up Obj. B5 in the process. We can see that this is not so, however, because B5 remains (and its relevance is even clarified) if we suppose that a suitable alternative theory can resolve B2 and B3 by giving identification a qualitatively distinct constitution. The problem captured in the regress and ab initio objections is distinct from the Arbitrariness objection to Frankfurt’s initial account of decisive identification, because the latter properly concerns the problem of ambiguity as Frankfurt describes it in "Identification and Wholeheartedness:" namely, as a conflict between identifications themselves. As Bransen notes, the difference between such ambiguity and the kind of conflict observed in the unwilling addict suggests that conflicts between identification-constituting states themselves are qualitatively different from those between an identification-constituting state and other states that can exist in us (and even move us to act) without our identifying with them at all. As a result, the issue which decisive identification resolves is different in kind from the issue which straightforward identification or identification per se resolves (e.g. in the unwilling addict’s case).

This in turn implies that primary regress objection (B2) cannot be answered simply by appeal to decisive identification, as Bransen suggests: just as modifications sufficient to resolve Obj.B2-B3 will not necessarily resolve B5, similarly, we cannot derive a solution to B2-B3 from an independent solution to B5. Yet in "Identification and Wholeheartedness,"

---

814 Frankfurt, "Identification and Wholeheartedness," p.175: "It may also turn out that there is a conflict within the authority [of the agent] itself — that the person has identified himself inconsistently. This is the issue of wholeheartedness."


816 Ibid, p.4.
after admitting the substance of Obj.B2-B3, Frankfurt unfortunately follows Watson’s lead and does not distinguish the primary regress problem relevant to the constitution of identification *per se* from the different regress problem which decisive identification is supposed to resolve. After admitting the inadequacy of his initial account of positive identification, Frankfurt *should* say, as I’ve argued, that the solution lies in the qualitative differences between *cares* and mere second-order desires. Instead, Frankfurt tries out the notion that this problem can be resolved by reformulating his initial account of *decisive* identification.  

It is easy to confuse Obj. B2 and B5 in this way, because it is easy to think that identification *per se* might be constituted by a simple second-order desire with which we decisively identify in the sense of being non-arbitrarily but unreservedly committed to it. The phrasing of Obj. B2 and B3 can suggest that the problem is (a) just to *add* something to second-order desires which will assure that they themselves are identified, and (b) that this *additional component* not lead to a regress of its own. Since decisive identification understood as wholeheartedness seems to fit the requirements for a solution understood *this* way (as opposed to my construal of the requirements in the Qualitative Condition), why not appeal to it? The problem with this move (and thus with the misconstrual of B2-B3 that motivates it) is that it is phenomenologically inadequate: for if we identify with a desire $D_1$ because we *decisively* identify with our desire $D_2$ that $D_1$ should be our will, then once again, ambiguity or a "conflict within the authority" of the self will be impossible, since "the person

---

817 This appears to be in part because Frankfurt hesitates to treat care as an analysis of identification, and because he follows Watson’s initial conflation of the issues concerning identification vs. decisive identification (see §3.2 - B5). He says: "I tried some time ago to deal with this problem in the following passage," and then quotes his 1971 paragraph on decisive identification ("Identification and Wholeheartedness", pp.166-167, my italics).
who has identified himself inconsistently" lacks wholeheartedness\textsuperscript{818} or does not decisively identify with any of his second-order desires, and hence on this proposed model would lack identifications entirely.

Yet what distinguishes ambiguity from wantonness is precisely that the ambiguous person still \textit{has identifications}, and hence is not \textit{completely} passive, although he may not be as absolutely active with respect to his motives\textsubscript{i} as is the person who decisively identifies with some of them. Wholehearted identification with a desire\textsubscript{i} means, as Frankfurt says, "that the desire is \textit{in the fullest sense} his, that it constitutes what he really wants."\textsuperscript{819} By contrast, the ambiguous person has no "unequivocal" answer to the question of what she "really wants" and is not \textit{fully integrated};\textsuperscript{820} she is moved to act "without wanting wholeheartedly to be motivated\textsubscript{[1]}" as she is, and hence she is "to \textit{some degree} passive with respect to the action [she] perform[s]" on this motive\textsubscript{i}.\textsuperscript{821} Note that all these phrasings leave room for the idea that the ambiguous person is \textit{also to some degree (though not ‘completely’) active} unlike the wanton or unwilling addict, as long as she acts on any of the motives\textsubscript{i} with which she has conflicting identifications.\textsuperscript{822} In short, wantonness, identification per se, and complete or wholehearted identification constitute three qualitatively \textit{distinct levels} of identification in general. The middle and highest levels cannot be collapsed without failing to adequately account for the diversity of identificatory phenomena in our inner lives.

\textsuperscript{818}Frankfurt, "Identification and Wholeheartedness," in \textit{The Importance of What We Care About}, p.175.
\textsuperscript{819}Ibid, p.170, my italics.
\textsuperscript{820}Ibid, p.165, my italics.
\textsuperscript{821}Ibid, p.163, my italics.
\textsuperscript{822}For note that it is also possible to be ambiguous in one’s identifications (identifying with conflicting motives, A and B, which both conflict with C) and still to act \textit{entirely unwillingly}, on motive\textsubscript{i} C, with which one does not identify at all. Indeed it is perhaps not uncommon that ambiguity in one’s (strongly) higher-order will may contribute to its inability to effectively control one’s actual first-order will\textsubscript{i}; wholeheartedness in identification may be \textit{contingently} linked to having an effective or "free" will\textsubscript{i} in Frankfurt’s sense.
IV.5.2. Decisive Identification as Volitional Necessity or Satisfaction

It still remains to consider Frankfurt’s new proposals as an account of how agents resolve ambiguity in decisive or complete identification. Understood in these terms, Frankfurt has actually presented two subtly different new analyses of "wholeheartedness:" in terms what he calls "volitional necessities," and in terms of an overall state of "satisfaction."

Volitional Necessity

The first appears in "Identification and Wholeheartedness," where Frankfurt argues that there are certain kinds of reasons that may justify stopping a potentially endless series of re-evaluations of a result at some point, such as that the person is "unequivocally confident" that it is the right answer, or believes he has good reason to discount whatever possibility remains that conflicting results could arise in light of further re-evaluations. In such cases, "resting content’ with the current result is not arbitrary, as Watson suggested:

Both in the case of desires and in the case of arithmetic a person can without arbitrariness terminate a potentially endless sequence of evaluations when he finds that there is no disturbing conflict, either between the results already obtained or between a result already obtained and one he might reasonably expect to obtain if the sequence were to continue. Terminating a sequence at that point—the point at which there is no conflict or doubt—is not arbitrary. For the only reason to continue the sequence would be to cope with an actual conflict or with the possibility that a conflict might occur.

Note that this implies that the reformulated account actually functions to address the separate problem of an actual or potential conflict in our very identification with motives,—the issue in Obj. B5, as I argued above. Frankfurt’s argument implies that the "wholehearted" person can completely or with "genuinely unreserved commitment"
identify with his existing positive identifications without any implicit higher-level re-evaluation, because these identifications themselves are not in conflict and there is no real reason to doubt their stability.

This account is brief, but it is connected in a way Frankfurt does not here elaborate with an earlier theme in "The importance of what we care about." The ‘possibility’ of conflict which is discounted in such non-arbitrary cessations of re-evaluation must be understood in the Cartesian sense as a real (more than just speculative) or substantiated one, rather than a mere logical possibility; otherwise we could never decisively stick with the current result of our evaluations unless contrary results were logically impossible — a condition virtually never met in practical deliberation. In fact, it seems that on Frankfurt’s first account, what decisive identification must rule out is a conflicting identification’s being volitionally possible for the agent. And this explain why Frankfurt also analyzes complete or wholehearted identifications in terms of "volitional necessities:"

There are occasions when a person realizes that what he cares about matters to him not merely so much, but in such a way, that it is impossible for him to forbear from a certain course of action. It was presumably on such an occasion, for example, that Luther made his famous declaration: "Here I stand; I can do no other." An encounter with necessity of this sort characteristically affects a person...by making it apparent that every apparent alternative...is unthinkable.826

It is important to realize that Luther may still have had rational access to other objective possibilities, such as that respect for tradition should lead him to remain silent. But volitional modality is synthetic, and an agent’s volitional possibilities will not include all options that are ‘possible’ in other senses, such as nomologically possible, or morally possible (deontically permissible). And as Frankfurt notes, "the impossibility to which

826Frankfurt, "The importance of what we care about," p.86.
Luther referred was a matter neither of logical nor causal necessity.\textsuperscript{827} Nagel suggests instead that "When Luther says he \textit{can} do nothing else he is referring to the normative irresistibility of his reasons, not to their causal power."\textsuperscript{828} But even if Luther does perceive his rebellion as \textit{normatively} necessary or as a \textit{duty}, this is not by itself a sufficient explanation of his action, since it is usually volitionally possible for a human agent to act in ways that would violate norms he recognizes: if he does not have such possibilities in a given case, it is because following this norm has become \textit{volitionally} necessary for him.\textsuperscript{829} Thus Frankfurt's insight is that there is such a thing as a distinctively volitional form of modality which is cannot be assimilated or reduced to any other modal type (including the moral modality recognized by practical reason). It is not that the person with a volitionally necessary care "cannot muster the \textit{power}" to act against it; rather "What he cannot muster is the \textit{will}."\textsuperscript{830}

Frankfurt argues that "volitional necessities" in this sense do not make agents "helpless bystanders to their own behavior," as the psychological necessity of the unwilling addict's desires do, for example; rather, the "person who is constrained by volitional necessity" might have it \textit{within his power} to act otherwise, but he is unwilling to do this, and "furthermore, his unwillingness is \textit{itself} something he is unwilling to alter."\textsuperscript{831} Far from being felt as alien restrictions, volitional necessities appear as \textit{fully identified} with the person and even "liberating" (in the sense of making the agent more autonomous or more

\textsuperscript{827}Ibid, p.86.
\textsuperscript{828}Nagel, \textit{The View From Nowhere}, p.116, footnote 3.
\textsuperscript{829}And, at least for human agents, this is not entailed by its being \textit{morally} impossible or impermissible.
\textsuperscript{830}Frankfurt, "On the Necessity of Ideals," p.20.
\textsuperscript{831}Frankfurt, "The importance of what we care about," p.87.
completely self-determining).\cite{ibid, p.89} As Frankfurt says, "People are generally quite far from considering that volitional necessity renders them helpless bystanders to their own behavior" in the way that, for example, the unwilling addict experiences. Rather, "they may even tend to regard it [the volitional necessity] as enhancing both their autonomy and strength of will."\cite{ibid, p.87} Thus when a person finds it "unthinkable" to act in some way or on a given motive, contrary to her volitional necessity, "the aversion is not only irresistible; it is also in some way endorsed by the person."\cite{Frankfurt, "On the Necessity of Ideals," p.21.}

There are some questionable assumptions made in Frankfurt’s analysis of volitional necessity to date. For example, it appears that by definition, a person could not have conflicting volitionally necessary cares, which is why this strong mode of caring can provide an explanation of complete or decisive identification. Also, Frankfurt seems to assume that volitionally necessary identifications always effectively control the first-order will, i.e. that they guarantee a modicum of free will in his sense. Yet if volitional necessity is a mode of identification, it seems worth asking whether it is possible for someone to be compelled to act on motives, with which he finds it volitionally impossible to identify, i.e. motives that are necessarily alienated from him.\cite{Frankfurt, "On the Necessity of Ideals," p.21.} In addition, Frankfurt has argued that to love and to

\footnotesize
\begin{itemize}
  \item \cite{ibid, p.89} \cite{ibid, p.87} \cite{Frankfurt, "On the Necessity of Ideals," p.21.}
  \item \cite{Frankfurt, "On the Necessity of Ideals," p.21.} One might question this claim by referring to evidence that persons exhibiting the phenomenology of a volitional necessity often ascribe it to an external forces or authorities, precisely because it is not under their voluntary control (e.g. Joan of Arc, who felt that God’s voice compelled her to her mission, or Nietzsche, who believed that an absolute fate had destined him to his role as destroyer of ‘slave morality’). But one could respond to this objection with the hypothesis that revolutionaries are often in bad faith about their volitional necessities, preferring to ascribe them to something beyond themselves, while in fact they identify with these supposedly external forces or authorities.
  \item \cite{While} While it may be hard to credit that an addiction could force someone to an act she finds absolutely “unthinkable” in Frankfurt’s sense, there may be psychological evidence to support the idea that in certain types of compulsive disorders and psychosis, a person may feel overcome by desires or intentions that horrify her, and which she could not normally even contemplate. This may indeed be one constitutive element in the feeling of “craziness” which attends this very unfortunate type of case.
\end{itemize}

351
have ideals require not merely states of care, but volitionally necessary caring:

We cannot help loving what we love, nor can we make ourselves love by a mere act of will. The self-fulfillment and freedom [i.e. autonomy] that love provides depends upon the very necessity that love entails. If someone loves nothing, it follows that he has no ideals. Now an ideal is a limit. A person’s ideals are concerns that he cannot bring himself to betray. They entail constraints that, for him, it is unthinkable to violate.836

These linkages present certain problems. First, there are other senses of "ideals:" a person may have ideals which he hopes are realized in the world, but which he has no opportunity to further or hinder in his own sphere of action. But if we suppose that there may be volitional ideals in Frankfurt’s sense, it seems that someone could have a quirky variety of things she could not bring herself to do (or even contemplate doing), without these cohering sufficiently to form an ideal. Also, we might have reason to think that at least some kinds of love are not experienced as unavoidable. For example, agape or "neighbor-love," while it is dispositional, may be sufficiently under our voluntary control that it can be morally expected of us, and few people ever manage to love in this way with the alternative being volitionally impossible for them.837

But these concerns aside, volitional necessities are necessities of identification, which Frankfurt understands as "to a certain extent self-imposed," because the person guides himself away from considerations that might tend to make his will more ambiguous. The person who is wholehearted through volitional necessity thus not only

...cares about following the particular course of action which he is constrained to


837And Frankfurt admits that there must be bases "on which a person can reasonably make a choice of ideals" (ibid, p.25). While it is possible to have irrational or even unconscious volitional ideals, it must also be possible for the person to link the determination of his ideals to some objective criteria of worthiness. And this means that they cannot simply be given as fixed features of our personal essence. Absolute givenness presents the danger of another type of arbitrariness, the opposite of that Frankfurt invokes volitional necessities to avoid.
follow. He also cares about caring about it. Therefore he guides himself away from being critically affected by anything—in the outside world or within himself—which might divert him or dissuade him either from following that course or from caring as much as he does about following it.\textsuperscript{838}

This account, however, has led Annette Baier to object that a necessarily wholehearted person would be a "fanatic," since he will not risk "scrutinizing" his cares.\textsuperscript{839} This seems to be a deep problem, since although it might be possible to be a wholehearted fanatic (unless we build extra normative conditions of rationality into decisive identification), wholehearted caring cannot be so "self-protective" that it necessarily bars attention, for example, to evidence that our care results from manipulation or false consciousness\textsuperscript{840}—for this would make it essentially fanatic. This shows that a version of Obj.B1.ii arises against the analysis of wholehearted or complete identification in terms of volitional necessities.

Decisive identification understood this way may not be arbitrary, but it is compatible with origins that seem to undermine its autonomy, and make it an expression of ideological extremism instead. The volitional necessity analysis faces another more subtle problem as a result of the fact that it apparently rules out weakness of will. Yet there seem to be cases where a person "decisively" or with the complete authority of their self embraces some care and its associated motives\textsubscript{1} M, but nevertheless still identifies positively (though not obviously not decisively) with acting on contrary motives\textsubscript{1} N. The agent’s decisive or

\textsuperscript{838}Ibid, p.87.

\textsuperscript{839}Annette C. Baier, "Caring about Caring: A Reply to Frankfurt," Synthese, Vol. 53, No.2 (November 1982): 273-290, p.273. Baier argues instead that authentic “caring about caring” must involve “asking whether what one cares about should be cared about,” or caring enough about one’s cares to care about whether they are objectively important (p.274), and she tries to find some criteria for this. But it is clear from the proposed role of volitional necessity in explaining how decisive identification can be non-arbitrary, however, why Frankfurt did not go this route. Yet Frankfurt himself admits (though half-jokingly) that the weakest and least salient of all human passions is “our love of the truth about ourselves” —Frankfurt, "The Faintest Passion," Proceedings and Addresses of the APA, Vol. 66, No. 3, p.5.

\textsuperscript{840}Ibid, p.285: "If genuine caring does and must protect itself against such knowledge of itself, then caring is not compatible with risky caring about caring, [and is] compatible only with regarding one’s caring as so important that one protects it against any threat, including threatening knowledge."
wholehearted identification with M-motives does not render N-desires merely alien; when she acts on them, her will is still free and she is not coerced, although her motivation is not what in the deepest sense she completely wants. If this is possible, then, although the decisive identification cannot itself be in any way ambiguous, it may still fail to control and unify all the agent’s identifications.

Consider someone who grew up in a ghetto neighborhood in Los Angeles and spent most of his youth with a gang, the ‘Trips,’ several of whose local members became like family to him. But through the determined efforts of a local church youth leader, he is persuaded to go to finish high school and go to college out of state. He eventually becomes a police detective, and returns to L.A. Though he is now, let us suppose, wholeheartedly committed to being a law-enforcer and cares unambiguously about fighting crime, one of his street sources offers him a chance for a sting operation that would destroy the Trips and get several of his old friends in trouble. He is ready to order the sting, but holds off at the last minute. He is not gripped by any volitional necessity here: he does not find destroying his old gang "unthinkable" in Frankfurt’s sense (after a week of regret and guilt, he takes hold of himself and orders the sting). But neither is his hesitation simply a result being assaulted by passions that are ‘external’ to his will, or that he finds alien to his self-commitments: rather he discovers that he still identifies with the emotions of gang loyalty. This is not a case of symmetrical ambiguity, however, since there is never any doubt about where our agent’s ultimate commitments lie. But against this remaining part of his inner character, his decisive

---

841See Frankfurt, "Rationality and the Unthinkable," in The Importance of What We Care About: "Sometimes, however, a person may be incapable of performing a certain action even though, so far as considerations of these kinds go, he is in a perfectly good position to perform it. It may be, further, that he thinks he should perform the action, and that he has a desire to perform it. If he is nevertheless unable to act because he finds the action unthinkable, then what keeps him from acting is a volitional matter. He cannot perform the action because he cannot will to perform it." (p.181).
identification with fighting crime takes some time to exert itself; for a period of ‘weakness,’ he acts on a less important but still undeniable care for his gang.842

The volitional necessity-analysis implies that contrary to appearances, our detective cannot be completely or wholeheartedly identified with being a policeman when he nevertheless acts on a contrary identification. In order for his identification to count as "decisive," by this account, he would have to have found not ordering the sting "unthinkable," even while he might have desired to avoid it (in other words, the emotions of gang loyalty could still be present in him, but only as forces that he wholly disowns). But this just seems to be too stringent a condition for decisive identification, or a settled identity: occasional deviation from one’s deepest character is possible without implying that this it is not one’s ownmost identity after all. Indeed, in the case described above, it might even seem inhuman for the detective not to experience at least some real hesitation and internal (not merely alien) doubts about destroying the gang that nurtured him. Decisive or wholehearted identification in human persons cannot require the same unmovability that it might for immortals or Valkyries (creatures representing pure will).

Volitional Satisfaction

These considerations suggest that volitional necessity is not required for decisive identification.843 However, Frankfurt has also proposed another analysis of complete or decisive identification in his 1991 APA Presidential Address, "The Faintest Passion,"

842 We might think of cases like this as presenting a special sort of weakness, i.e. ‘weakness of the higher-order will.’ Or it might turn out that all genuine cases of weakness of will have something akin to this structure (but this question raises issues that I cannot now pursue in the detail they deserve).

843 They do not, however, challenge the view that volitional necessities are possible for human persons. In this respect, we might argue (in an Aristotelian vein) that to constitute decisive identifications, what are now "volitional necessities" for the agent must have arisen in part through a voluntary control involving alternate possibilities. I will return to this issue in the projected additional chapter on free will.
beginning with a new description of ambiguous or ambivalent "volitional movements or tendencies:"

First, they are inherently and hence unavoidably opposed; that is, they do not just happen to conflict on account of contingent circumstances. Second, they are both wholly internal to a person’s will rather than alien; that is, he is not passive with respect to them....

[By contrast] Conflicts involving first-order psychic elements alone —for instance, between attraction and aversion to the same object— do not pertain to the will at all. They are not volitional, but merely impulsive or sentimental. Conflicts that pertain to the will arise out of a person’s higher-order, reflective attitudes.\textsuperscript{844}

The first condition reflects the fact that each of the opposed states involves a strong evaluation of motives, and hence they conflict essentially. The second distinguishes ambiguity from the conditions of wantonness and alienated first-order desires.\textsuperscript{845}

But if complete or decisive identification is simply wholeheartedness in this sense, then it only requires that a person (a) have identifications and (b) lack ambiguity. In that case, for example, an unwilling addict decisively identifies with his healthy motive\textsubscript{1} simply because he does not also have the volition\textsubscript{2} of a willing addict: "The unwilling addict is wholeheartedly on one side of the conflict from which he suffers, and not at all on the other."\textsuperscript{846} Yet it seems that it could just be by accident that the addict is only unwilling and not ambiguous about his desire for the drug. Thus on the volitional necessity model, simply being unopposed was not sufficient to make a positive identification complete or decisive, since, e.g., it could remain volitionally possible for the unwilling addict also to form the higher-order volition of a willing addict, and become ambiguous. On the earlier account, then, to identify with a motive\textsubscript{1} without also identifying with it completely or ‘in the highest

\textsuperscript{844}Ibid, p.8. Note how radically Frankfurt’s terminology has changed from his "Freedom of the Will" paper 20 years earlier, where he used the term "will" for the effective first-order desire. Now the terms "will" and "volition" are reserved for states that have the internal quality or authoritative status of identifications.

\textsuperscript{845}The unwilling addict, as Frankfurt explicitly says, is "not volitionally divided or ambivalent" (p.9).

\textsuperscript{846}Ibid, p.9.
degree’ did not require actual ambivalence, but only the possibility of it.

On the new account, instead, Frankfurt defines “wholeheartedness” as the simple absence of ambivalence: "A person is volitionally robust when he is wholehearted in his higher-order attitudes and inclinations" (or dispositions of care). This is compatible with continuing psychic conflict, as long as the person is "resolutely on the side of one of the forces struggling within him..."847 In this sense, as Joel Anderson has suggested, wholeheartedness is conceived as "a certain degree of integration and equilibrium within the volitional makeup of the person."848 Frankfurt explains this equilibrium in terms the notion of volitional satisfaction: an agent’s being wholehearted with respect to psychic elements like beliefs, feelings, attitudes, or intentions, to act in a certain way "consists in his being fully satisfied that they, rather than others that inherently (i.e. non-contingently) conflict with them, should be among the causes and considerations that determine his cognitive, affective, attitudinal, and behavioral processes."849 Satisfaction in the sense meant here is not a maximizing concept: one can be satisfied with more than one element of one’s outer character (or first-order motives) while considering one more "urgent or compelling;" similarly, being satisfied with an element does not mean judging that it is "the best condition available to him," but only that it is good enough.850 Thus decisively or wholeheartedly identifying with an emotional disposition, for example, is not equivalent to preferring it


848Anderson, “Disputing Autonomy: Structural Hierarchicalism, Procedural Hierarchicalism, and the Grounds for Ascription,” presented at the Central Division Meeting of the APA (Chicago, April 1995), revised version on file with the author, p.5. However, following the same confusion I have diagnosed which conflates identification per se with decisive identification, Anderson treats Frankfurt’s condition of structural equilibrium as the property that needs to be added to higher-order volitions to endow them with the authority of identification at all.


850Ibid.
most out of the range of dispositions one could cultivate (which would entail being hostile to any possibility of change in this state once attained). This much of Frankfurt’s satisfaction-analysis is in keeping with his earlier arguments in "Identification and Externality:" just as to identify with a motive, is not to rank it highly in a single scale of higher-order preference, identifying with it decisively is not to rank it highest on such a scale. What is new, however, is the idea that the decisive identification is primarily a negative concept. For the satisfaction in terms of which wholeheartedness is now interpreted entails mainly

...an absence of restlessness or resistance. A satisfied person might willingly accept a change in his condition, but he has no active interest in bringing about a change. Even if he recognizes that he could be better off, the possibility does not engage his concern.851

This confirms that being satisfied with some first-order state S is a weaker condition its being volitionally necessary to care about S: the latter entails the former, but not the converse.

But as with the analysis in terms of volitional necessity, the satisfaction-analysis aims to provide an alternative to conceiving full or complete identification either in terms of a regress of ever-higher endorsements or an arbitrarily ‘highest’ endorsement. "Satisfaction" avoids this problem through not being a component within the structure of the agent’s will, but rather an overall condition of that structure. The person does not have to decide to be satisfied, or judge that he is: there is "no danger here of a problematic regress" because

Satisfaction with one’s self requires...no adoption of any cognitive, attitudinal, affective, or intentional stance. It does not require the performance of a particular

851Ibid, p.12-13. This idea is obviously related to Frankfurt’s "concept of having enough" in "Equality as a Moral Ideal," reprinted in The Importance of What We Care About: 134-158, p.137. In this essay, Frankfurt deployed this idea in arguing against Rawls’s use of open-ended primary goods in his theory of distributive justice, in favor of something like a distribution maximizing the satisfaction of basic needs (p.144), but he did not relate this concept to identification.
Contrast this with the analysis in "Identification and Wholeheartedness," which construed decisive identification as constituted by deliberately refraining from further re-evaluation on grounds that conflict with the current state was either impossible, or its residual possibility was irrelevant (grounds that volitional necessities would provide: e.g. Luther could deliberately stop re-evaluating his identification with the motives that moved him to post his Theses).

This is not quite the whole picture, however, because in order to assure that wantons cannot be "satisfied" (and thus fully identified with) their wills, by default, Frankfurt must add that the relevant absence of interests in changing one’s first-order states — i.e. the ‘equilibrium’ — must itself be motivated by the agent’s "understanding and evaluation of how things are with him." Wholehearted satisfaction thus still involves higher-order states that reflectively evaluate one’s first-order motives, emotion, dispositions, etc. It is these evaluations themselves with which the person must be ultimately satisfied in wholehearted identification.

**Three Problems with the Satisfaction-Analysis**

There are three problems with this account as it stands (in increasing order of seriousness). (1) The first is not a problem with satisfaction as an analysis of wholeheartedness, but rather with the fact that (following the same conflation in "Identification and Wholeheartedness") Frankfurt treats this analysis of wholeheartedness as solving the regress and ab initio problems for identification per se. After noting that being (weakly) second-order gives a desire no special authority, he writes:

---

852 Contrast this with the analysis in "Identification and Wholeheartedness," which construed decisive identification as constituted by deliberately refraining from further re-evaluation on grounds that conflict with the current state was either impossible, or its residual possibility was irrelevant (grounds that volitional necessities would provide: e.g. Luther could deliberately stop re-evaluating his identification with the motives that moved him to post his Theses).


854 Ibid. And thus, since wholeheartedness and ambivalence are only "attributable to persons," not to wantons, "wholeheartedness is not exactly equivalent to the absence of ambivalence," since it also involves the condition that one have some "reflective self-evaluation" or identifications to begin with (note 14, p.16).

855 Ibid, p.14. Yet Frankfurt still calls them "endorsing higher-order desire[s]," rather than using some other term that reflects their qualitative difference from mere preferences.
Hierarchical accounts of the identity of the self do not presume, however, that a person’s identification with some desire consists simply in the fact that he \textit{has} a higher-order desire by which the first is endorsed....Identification is constituted neatly by an endorsing higher-order desire with which the person is satisfied.

But, aside from the fact that this construal seems to lose what was gained earlier in the care analysis,\textsuperscript{856} it would be impossible on this account to distinguish ambiguity from wantonness, since the person lacking in complete or wholehearted identification would appear to be lacking in identification \textit{altogether}. This is just a repeat of the problem diagnosed above (§5.1).

(2) Suppose, then, that we assume "wholeheartedness" stands only for complete or decisive satisfaction, as opposed to identification \textit{per se} (which is compatible with ambiguity), and regard the satisfaction-analysis solely as an interpretation of wholeheartedness in this distinct sense. Then it is still inadequate, because, as already pointed out, it has the implication that someone who identifies with a motive, $M$\textsuperscript{857} and \textit{happens} to lack any contrary cares, is for that reason \textit{satisfied} with $M$ and thus completely or wholeheartedly identified with $M$. This implication is counterintuitive, since an identification that would still exist if the agent became ambiguous is not more complete or qualitatively higher as an authoritative expression of the self \textit{merely} for lacking a competitor. Although it would be wrong, for example, to call the willing addict’s simple unopposed positive identification with his addictive desires, ‘\textit{halfhearted},’ because it is after all unopposed, the identification still need only be ‘with heart,’ not necessarily ‘with all of his heart.’ We can suppose that he identifies with his cravings without supposing that he cares

\textsuperscript{856}Which added temporal extendedness and decisions about types of motives, based on strong evaluations to the conditions of identification, as we saw.

\textsuperscript{857}Through caring about something that requires a strong positive evaluation of motives like $M$ and a decision to endorse $M$ on this basis.
about maintaining this lifestyle in the deepest possible way in which persons can be volitionally committed to something. In short, *ambiguity* or ‘halfheartedness’ and *wholeheartedness* (in the sense of complete identification or maximally authoritative devotion) are not jointly exhaustive disjuncts: nonambiguous but still nondecisive positive identification is a third possible alternative.

(3) Finally, unlike the interpretation of wholeheartedness in terms of volitional necessity, the satisfaction-analysis *in principle* leaves open the possibility that one might completely identify with being in ambiguity. If one had essentially conflicting higher-order volitions endorsing opposed first-order dispositions, then one’s reflective evaluations of one’s states would be inconsistent, but suppose one lacked any inclination to resolve the conflict *despite* this inconsistency. Though it would be difficult to apply the label ‘wholeheartedness’ to this condition, it *would* nonetheless constitute decisive identification with an ambiguous will, according to the satisfaction-analysis. Anticipating this problem, Frankfurt argues that

A person may come to accept the fact that he is ambivalent as unalterable. It seems to me, however, that is not a fact with which he can possibly be satisfied. No one can be wholeheartedly ambivalent, anymore than someone can desire unequivocally to betray himself or to be irrational…It is a necessary truth about us, then, that we wholeheartedly desire to be wholehearted.\(^{858}\)

It is a very interesting and highly important question whether this is true. If it is, then *some* kind of self-integration remains an *absolute telos* of persons (in the sense that it is built into the natural structure of personhood that they cannot deviate from this end). As suggested in my Introduction, there may be phenomenological evidence against this which would show that even if wholeheartedness is indispensable for a *fully meaningful life*, it can still be

---

\(^{858}\)Ibid, p.14.
rejected.\textsuperscript{859}

But there are also different ways one could take Frankfurt’s thesis. (i) If it were interpreted only as saying that persons necessarily identify in the complete or fullest way with the project of determining some commitments or cares that have decisive importance for them, or that can be absolutely authoritative expressions of their ‘innermost’ will, the thesis would be compatible with existential conceptions of personhood,\textsuperscript{860} and be supported, for example, by Bernard Williams’s argument that we must have ”ground projects” that we would die for to have any persistent self at all.\textsuperscript{861} But on this reading, the thesis would make no reference to the concept of satisfaction, and would leave open the possibility that wholehearted or complete identification is something other than being ”satisfied” with our first-order states.\textsuperscript{862}

(ii) If instead the thesis is that we must completely or decisively identify with being wholehearted in our cares.

\textsuperscript{859}For instance, in Kierkegaard’s Sickness Unto Death, tr. Howard V. Hong and Edna H. Hong (Princeton University Press, 1980), describes a virulent form of despair as one in which the agent tries to form his ”concrete self” (or outward character) entirely by sheer fiat, and ends up identifying only with its complete ambiguity (or lack of irreversible commitment to any concrete features): yet he \textit{wills} just this. In defiance, he also identifies with any concrete feature that he could not change even if he wanted to: ”Rather than to seek help, he prefers, if necessary, to be himself with all the agonies of hell” (pp.68-72). This can lead to the most malicious and intense kind of despair of all, which Kierkegaard calls ”demonic” despair (pp.72-73).

\textsuperscript{860}For instance, Heidegger seems to hold that an individual Dasein is always committed at any given point to some ultimate end or end which provides the unity of the ”totality of involvements” in terms of which the significance of everything in his ”world” is first given to him: ”In understanding a context of relations such as we have mentioned, Dasein has assigned itself to an ‘in-order-to,’ and it has done so in terms of a potentiality-for-Being for the sake of which it itself is—one which it may have seized upon either explicitly or tacitly, and which may be either authentic or inauthentic” (\textit{Being and Time}, I.3, H86, p.119). Thus even if my ultimate for-the-sake-of-which is passively taken over from one’s society, or is unclear and conflicted, I can never be without one.

\textsuperscript{861}Williams, ”Persons, character, and morality,” in \textit{Moral Luck}, p.12-13.

\textsuperscript{862}One reason for taking the thesis this way is that it seems to be a stronger version of an earlier idea which Frankfurt expressed without the notion of satisfaction as follows: ”...it is necessarily important to people what they care about,” since their caring always makes ”an important difference” to them, from which it follows that ”if anything is worth caring about, then it must be worth caring about what to care about” (”The importance of what we care about,” p.92). This could be rephrased more clearly as follows: if it is worthwhile not to be \textit{wanton}, or to form identifications at all, then it is volitionally necessary to care about \textit{which} identifications we form, and why. This thesis, which I find plausible, is however notably weaker than the thesis that if we are not \textit{wanton}, it is volitionally necessary to care about being \textit{wholehearted} in our cares.
satisfied (in Frankfurt’s negative sense) with our volitions and related evaluations of our psychic states, then it still seems plausible, but only because this formulation does not rule out being satisfied with conflicting or ambivalent higher-order evaluative volitions. And again, this would leave open the possibility that one could wholeheartedly identify with being unsatisfied—even if one could never, strictly speaking, be satisfied with being unsatisfied.

(iii) Instead, Frankfurt appears to mean that the ambivalent person always has the volition, not to be ambivalent and is satisfied with this volition, and could never be satisfied with the opposite volition, should it occur. This seems far less plausible—which may simply suggest that wholehearted or complete identification is not constituted by such satisfaction (in which case version (i) could remain true). But even if version (iii) were true, this fact would not follow just from the analysis of decisive or wholehearted identification in terms of satisfaction: (iii) remains an additional premise. And this means that on the satisfaction-analysis, wholeheartedness is negatively interpreted in such a way that it does not by definition rule out ambivalence. Yet this was the original condition for something’s counting as decisive or wholehearted identification: the idea was that the nature of the highest kind of authority would show why ambivalent identification could not be a completely authoritative expression of the self. Following the Qualitative Condition, it is part of their constitution that we are already completely identified with our decisive identifications or most authoritative ‘internalizations’ of other states. The satisfaction analysis fails to meet this crucial condition.

Conclusion

It seems, then, that decisive identification must involve conditions stronger than mere satisfaction, but probably need not involve conditions as strong as volitional necessity. The
error in both Frankfurt’s more recent strategies for analyzing full or decisive identification seems to be the assumption that the complete or highest degree of identification is constituted by ordinary identification plus some additional property. In the first strategy, the extra property is a modal one: decisive identification is just identification per se that is volitionally necessary for the agent. In the second strategy, the extra feature is weaker, nonmodal, and negative: identification per se becomes decisive if the agent lacks any conflicting identification or actual tendency to question or doubt his existing cares. In both cases, however, complete or decisive identification is conceived just as some reflexive relation of ordinary, positive identification to itself. This seems to me analogous to the error of Frankfurt’s initial model in trying to explain identification per se simply in terms of a reflexive relation of desires: in parallel, decisive identification is now constituted either by caring about our cares in a way that makes them necessary for us, or by an "absence of discord" between our cares, which makes us satisfied with them. These analyses of decisive identification succumb to objections analogous to those against Frankfurt’s initial model of identification per se because they fail to recognize that complete or highest-order identification can be constituted only by states qualitatively distinct from those involved in identification per se, which are thus even further removed from mere preferences or S2-desires. To explain wholehearted or complete identification, we need to look for a different kind of state, not merely a modification of or relation among our positive identifications.

IV.6. The Projective Account of Identification

IV.6.1. Three Relevant Orders of Identification and the ‘Personal World’

364

363In the strong or thick sense of ‘order’ as authority for the self.
At the phenomenological level, then, it has become clear that we need to recognize at least three qualitatively different types or levels of ‘identification’ itself, none of which are constituted simply by variations or combinations of the states constituting the lower type. For just as the experience of identifying with desires and emotions that motivate first-order acts is qualitatively distinct from having preferences or nonevaluative desires (S2-inclinations), the experience of completely or decisively identifying with our passions and motives, involves yet a further qualitative distinction, since we may genuinely identify with a desire, and not be wanton with respect to it, without this guaranteeing that our identification is complete and not subject to ambiguity. And, as Gary Watson mentions, there is a weaker sense still in which every first-order desire or emotion is recognized as ‘ours.’ At the phenomenological stage of analysis, then, it is helpful to think of these as different degrees or levels of identification which form a ‘hierarchy’ in the strong or thick sense of representing qualitatively different orders of authority or inwardness in relation to the self:

- Minimal identification (Identification$_1$)
- Positive (or per se) identification (Identification$_2$)
- Decisive (or complete) identification (Identification$_3$)

These, then, are the main "orders of volition" relevant for understanding autonomy and authentic selfhood. The phenomenological differentiation of these "orders" as distinct levels of agent-authority is apparent from the following differences, summarized in this table:

<table>
<thead>
<tr>
<th>Order of Identification</th>
<th>Status</th>
<th>Unity</th>
</tr>
</thead>
</table>

---

864 This framework could be further elaborated. For example, as I suggested earlier, between mere consciousness of first-order motives (identification$_1$) and explicit or deliberate identification with them (identification$_2$), we might also discern a level of tacit identification; however, since the dispositional account of identification in the care-analysis may be sufficient to explain some types of tacit identification, I will not include this as a separate level here.
However, it is possible to argue that this minimal identification does not follow simply from consciousness, or that it involves something more than states being among our actual or possible conscious states. But this argument would take me too far afield here.

This description is meant to be neutral between divergent accounts of motivation. I have not said "desires of which the agent approves," or "desires that the agent values as objectively good to be moved by," or "desires the agent wills to be moved by," since any of these phrasings would lean towards one of the constitutive theories between which we have yet to decide.

Our preliminary phenomenology of identification discloses these different qualitative orders of intrapersonal self-relatedness prior to the introduction of theoretical hypotheses to explain how these fundamental types of intrapersonal relation are constituted. This is the main result of the qualitative approach to the phenomenology of identification: it is the orders of identification, not mere iterations of desire, which are important to moral personhood. Let me develop this important result by describing the three main orders of identification in more detail.

(A): Identification, As Modally Extended

We are minimally identified, with desires, emotions, intentions, and cognitive states of which we are conscious. They are ‘ours’ in a weaker sense than in identification per se (identification), i.e. identification with desires or preferences (either simple or complex) that the agent internalizes or recognizes as motives by which she would be moved. Positive identification in this sense, however, presupposes that we are ‘minimally identified’ (or identified) both with those actual first-order motives of which we are conscious, and with a larger extended sphere of first-order states (e.g. emotions or desire complexes) that are psychologically possible for us. Whether we adopt them as our own, alienate them, or are wanton with respect to them, first-order motives that are either operative in our actions, or

---

865 However, it is possible to argue that this minimal identification does not follow simply from consciousness, or that it involves something more than states, being among our actual or possible conscious states. But this argument would take me too far afield here.

866 This description is meant to be neutral between divergent accounts of motivation. I have not said “desires of which the agent approves,” or “desires that the agent values as objectively good to be moved by,” or “desires the agent wills to be moved by,” since any of these phrasings would lean towards one of the constitutive theories between which we have yet to decide.
have the potential to become so, already have one sort of irreducibly first-person significance, i.e. a meaning independent of any shared, third-person characterization of them.

This involves one essential modification of Frankfurt’s descriptions. In his "Freedom of the Will" paper, Frankfurt suggested that if agent A has a volition \( z \) to X, this "entail[s] that A already has a desire to X," although this desire may not necessarily be the will \( w \) on which he acts.\(^{867}\) Call this the Actualist Restriction (AR): we must actually have a desire, however weakly, in order to identify with it. Frankfurt thinks the only exceptions to AR would be "non-standard" cases where someone wants to have a will \( w \) that fits some condition with unbound variables, such as "whatever desire effectively moves [another person] B," though A does not know in advance what this might be.\(^{868}\) But AR is implausible on its face. For one thing, it has the implausible consequence that a mere second-order desire to experience some first-order desire could never be wholly unsatisfied, since it could never even exist unless the desire \( d \) it was about already existed in the person. More importantly, however, it is counter to experience to insist, for example, that an unwilling addict must have experienced the desire \( d \) not to take drugs in order to identify herself with it. On the contrary, her addictive desires may be so powerful as to exclude even a feeble desire not to take the drugs (or a desire to remain in a chemically unaltered state); yet surely she can still identify with this counterfactual desire \( d \). We can, then, identify with desires, emotions, or other first-order motivational states that are only ideal for us, even though we do not now (and maybe never have) actually experienced them. The range of identification \( z \) is modally extended to states that are psychologically possible for the agent. Thus minimal identification \( w \) must also extend to this same range of states. This does not mean that we can just identify with any

\(^{867}\)Frankfurt, "Freedom of the Will and the Concept of a Person," p.15.

\(^{868}\)Ibid, note 4, p.15.
state whatsoever; the state must be psychologically possible for us in the sense that (a) we have some idea of what it would be like to experience that state (which includes that it is biologically possible for a being of our species to attain this state), and (b) that it is psychologically possible for beings like us to will it.\textsuperscript{869}

Frankfurt’s attraction to AR could be a result of the more emotivist tenor of his early work, which attempted to keep identification consistent with a minimalist or Humean moral psychology. Alternatively, it could be motivated by the assumption that identification only becomes an issue when there is some conflict between the states with which the agent might identify. For example, Frankfurt says that the "unwilling addict has conflicting first-order desires" between which the higher-order volition chooses: "he wants to take the drug, and also wants to refrain from taking it."\textsuperscript{870} But in fact, identification requires not an actual conflict but only a potential conflict. We can calmly and wholeheartedly identify with the only desire, we happen to experience to any notable degree in some action-circumstance, because it still conflicts (counterfactually) with other desires that are psychologically possible for us.

A person thus has something like a modally extended horizon of actual and psychologically possible first-order states with which she could identify. All of these states are identified, in the most elementary sense with the agent: she owns them in the minimal sense that they are qualified with the first-personal significance of being psychologically possible states for her. In this respect, minimal identification forms something like a first-order world of the person, a realm of psychologically salient potentia.

\textsuperscript{869}Note that it is probably not ever psychologically impossible to will a state we actually have—even if it might be volitionally impossible for us to identify with it. But this condition would rule out, for example, states that no human person could identify with, such as being in prolonged and terrible agony.

\textsuperscript{870}Ibid, p.17. Thus this unwilling addict satisfies AR.
(B): Identification: Penelhum and Frankfurt

Positive identification refers to an achievement of personal self-relation which is qualitatively distinct from minimal identification, or the basic first-personal significance which qualifies psychologically possible and actual states. Moreover, positive identification presupposes minimal identification, but not the reverse. To identify with certain desires and preferences—or to will that they become part of the outward character I will to be—entails willing that I acquire these desires and preferences as elements of my own mental experience, if I do not have them already. I cannot authentically identify with your operative motives, for example, without willing to experience the same motivations myself. But as the unwilling addict’s experience testifies, I can certainly discover as mine in the ‘minimal sense’ operative desires, impulses, and motives which I emphatically do not will to be part of my outward self.

This dependence of identification on the ‘minimal’ first-personal significance of one’s own (actual and possible) preferences and affective states is implicit in Frankfurt’s own arguments against Terence Penelhum in "Identification and Externality." In his 1971 article, "The Importance of Self-Identity," Penelhum agrees with Frankfurt that as opposed to the "numerical" identity of a person, which may involve no more than the Lockian requirement that "B at the later time have memories of A’s actions or experiences at the earlier time," "self-identity" requires that a person has "integrated" himself such that "the desires that move him to act are desires that he acknowledges and wishes to be moved by."
But when someone cannot achieve this "self-integration," one strategy for overcoming the problem is to conceal it. And one method of concealing it is to "so describe the act as to represent the desire, and perhaps also its execution, as external to himself." Penelhum protests that this form of bad faith is "moral trickery" which some deceptive philosophical theories, such as "the Gallic concept of the crime passionel" (but not Frankfurt’s theory) too easily permit:

To say harmlessly that one is governed by a desire that is not one’s own is to utter a metaphor the literal translation of which is that one is governed by a desire that one does not want to be governed by. To say that the desire is not one’s own and mean this literally is to say something obviously false: for the desire is operative and therefore exists, and is not someone else’s.

Frankfurt responds by defending the notion that a desire may be "external" to a "person," i.e. to his second-order will. Using the analogy of a forced bodily impulse, Frankfurt argues that a "human desire" may not be the desire of any particular "person."

..we find it useful to reserve a sense in which a movement of this kind is strictly attributable not to the person at all, but only to his body...Now why may a desire not, in a similar way, be an event in the history of a person’s mind without being that person’s desire? Why may not certain mental movements, like certain movements of human bodies, in this sense belong to no one?

Frankfurt’s point is that there is a sense of identification in which a desire, just like an exterior force, may not be ‘identified’ in the volitional sense with the person it affects. Only second-order volitions that aim at identification in this more substantial sense establish his inward or personal identity. The person in Frankfurt’s sense is this ‘inward self’ that embodies the will to be a certain sort of ‘outward self.’

---

874(...continued)
Penelhum would seem to indicate.
876Ibid.
877Frankfurt, "Identification and Externality," p.61, my italics.
878I doubt, however, that Penelhum would disagree with this. He admits that the capacity to desire (continued...)
Nevertheless, Frankfurt’s analogy between "external" bodily movements and states of mind does leave out an important existential feature of personhood—a feature which Penelhum is trying to preserve. Movements that are part of my body’s history may not even be consciously recognized as mine at all, but at least all my states of conscious experience do come with a minimal identification of first-person ‘mineness.’ Even conscious impulses that I find alien to me in Frankfurt’s sense, because I authentically identify with a volition not to act on these impulses, are still alien only in an intrapersonal sense. Using my subscripts, we can say that they are impulses of mine, with which I do not identify.

As with identification, then, we must distinguish orders of ‘alienation.’ An experience may belong to another person rather than to me, since he (rather than I) may anticipate it as a possible experience for him. Such an experience seems alien, to me when I not only do not share it, but even feel that it not psychologically possible for me: e.g. I cannot possibly experience what it was like to participate in Celtic human sacrifice. It is quite another thing, however, when I experience desires and motivations that are (actually or potentially) part of my psychic life, and find them alien to me. Alienation in this sense—the inverse of positive identification—does not literally mean ‘belonging to an other,’ a wholly different

\[\text{\textsuperscript{879}(...continued)}\]

\[\text{\textsuperscript{879}}\] The modal sense of this impossibility, of course, is not logical modality, or even nomological. Rather, it refers to possible experience for the individual, or (as I put it) to psychological possibility.
person. This impulse I have alienated must still be minimally identified with me. I cannot alienate it in this sense if it lacks all first-personal significance or is psychologically alien for me.

Thus Frankfurt is right that "there is in fact a legitimate and interesting sense in which a person may experience a passion that is external to him and that is strictly attributable neither to him nor to anyone else:880 this is alienation, a volitional alienation that distinguishes between parts of one’s affective character that are all identified with the person as part of her conscious or unconscious life. When seen in light of these distinctions, however, what is phenomenologically most interesting is that in our common experience, the sort of identification involving higher-order volitions stands out against minimal identification with a desire, sentiment, or impulse, i.e. its being an actual or possible part of one’s own mental life. Qualitative differentiation of several types of first-personal significance is an essential feature of the life of persons.

The range of a person’s volitionally possible identifications thus constitutes something like inward horizon, or authoritative personal world of the self. This (strongly) higher-order ‘personal world’ is narrower in scope than the personal world of minimal identification, and it refers essentially not only to the person as owner of her conscious or unconscious mental life, but also to the person as volitional self, as a source of identifications. Since a person’s ‘inward character’ reflects not only her actual identifications, but also the range and relative availability of alternate identifications that are volitionally possible for her, the ‘shape’ of this inner personal world is an index of this inner character. If she is more inwardly disposed to certain types of cares than others, the contrary identifications may remain

880Frankfurt, "Identification and Externality," p.61; my italics.
volitionally *possible* for her, but they will be *less* willable than the identifications that best fit the cares to which she is disposed. Although Frankfurt does not say this, volitional possibility is *synthetic* not only in the sense that it is narrower in scope than logical possibility, but also in the sense that (unlike logical possibility) it *comes in degrees*, some possibilities being *more accessible* to the agent than others. The higher personal world of the inward self should therefore not be pictured as a circular ‘halo’ of possibilities surrounding the agent’s actual cares and commitments, but rather as an irregular ‘cloud’ of potential identifications, whose direction from the center reflects the aspect or domain of life to which they relate, and whose distance from the center reflects their relative degree of choosability for the agent. Beyond the edges of this ‘cloud’ may be some options that are *completely* impossible (in the volitional sense) for the agent. But it is not yet clear that it must have any very definite edge: note that lacking a sharp boundary would not prevent one’s sphere of volitional possibility from having a definite center.881

(C): *Identification,* 3: The Highest-Order Will:

In the context of Frankfurt’s initial hierarchical account, in principle there could be volitions of any order. But the qualitative version of hierarchicalism I have advocated mandates that in the thick or authoritative sense of ‘order,’ decisive identification represents a kind of *maximal* or *complete* type of identification. It is therefore appropriate to speak of a person’s decisive or wholehearted identifications as forming her *highest-order will,* the most inward core of her ‘self.’ If we cannot analyze decisive or wholehearted cares either in terms of volitional necessity or satisfaction, as argued in the previous section, then it may be plausible to think that cares with which a person completely identifies involve a decision

881 This model fits the notion of constrained liberty to be developed in the projected additional chapter on free will.
that is third-order (in the iterative sense), namely a decision to identify with one’s existing identifications resolutely and without reserve. But to constitute one’s highest-order will, volitions, of this kind would have to be effective in giving certain cares a kind of *maximal importance* for the agent, and presumably also give them a heightened sort of resilience—though perhaps not to the point of making their continuance absolutely necessary for the agent. In that case, *highest-order* identifications (in the authoritative sense) would typically involve (weakly or iteratively) *third-order* states of some kind, such as evaluations and decisions *about* our cares.882

I would like to suggest three points that would be central to a fully developed existentialist account of the maximally inward level of the self. First, a person’s highest-order will must, by its very nature, be unambiguously his. Ambiguity at this level is constitutively impossible.883 We might say that the activity of forming volitions of this kind is not simply predicated of the agent, but in an existential sense literally *is* what he is. For this reason, I refer to this level as the *existential will*. The possibility that the highest-order will is constituted by cares involving *third-order* attitudes and dispositions of the right quality and formed in the right way may also help explain other phenomena. For example, the Kierkegaardian idea that to leave wanton aestheticism, persons must inwardly choose to be makers of significant choices, or commit to becoming ‘selves’ with cares and commitments defining an inward identity for themselves, suggests that at the highest level, persons must either identify, with forming identifications, or identify, with remaining

882Note that Frankfurt’s own discussion of *caring about* our cares implies the existence of third-order states in this sense (see "The importance of what we care about," p.87 and pp.92-93).

883Thus if a person appears to have divided highest-order commitments, instead she is really identified, with remaining undecided or not fully committed in her cares.
Similarly, an adequate analysis of "bad faith" would show that this phenomena, as Sartre describes it, must also involve complete commitment to certain third-order attitudes.  

Second, as these examples suggest, it is plausible to think that the higher or more inward a volitional disposition is, the less likely its content is to be immediately transparent to the agent, or accurately assessed by him even in self-reflection. To this extent, the common metaphors of depth and height may be appropriate. We may discover what our highest or most inward commitments are only in crises, or circumstances when conflicts arise between our caring about different things to which we have been devoted. For the most part, our highest-order will may remain tacit, already affecting our view of the world and shaping our perceptions of options and what is important in life, yet without having been explicitly faced or shaped by our conscious decisions at all. Accordingly, a person may have highest-order identifications without being aware of them in the course of everyday life. Yet it also seems likely that when they do rise to the occasion and face this question explicitly and clearly —even if their self-assessments are not wholly accurate— persons can have some deliberate control over what they care about most deeply. 

Third, despite the impression given by Frankfurt’s initial account, it may also be impossible not to have any highest-order or completely authoritative commitments —even if it is only a tacit identification, with not facing the question. This is the real significance of
the existentialist dictum that ‘not deciding is still a choice.’ The complete or decisive commitments that a person has simply by virtue of the unavoidability of having some highest-order will, may therefore be woefully insufficient to integrate all her cares and motives, or bring substantial unity to her entire motivational structure, but they show that within her personal world, the person herself cannot completely avoid standing for something, however minimal, which a perfectly self-knowing being could equate with her innermost self. There is always at least a germinal core of a self within even the most disordered and divided volitional hierarchy.

This is no more than the barest sketch of how an existentialist account of the highest-order will or decisive identification, would look, but it indicates sufficiently for my present purposes how an existential analysis can pick up the threads of Frankfurt’s account where he left off. Since pursuing this development in detail would require substantial exegesis of some major existentialist writings, however, this task must be left for a separate occasion. Instead, I will argue that the analysis of projective motivation (Chap. I - III) provides the basis for an account of positive identification, that deepens the advances Frankfurt has made in his conception.

IV.6.2. Identification as a Distinctively Volitional Type of Motivational State

In this chapter, we have seen that Frankfurt’s best prospect for resolving the main objections to his initial model lay in construing identification as constituted by caring about something, which is a reflexive structure including persisting dispositions (a) to give strong positive evaluations of motives, typically involved in concern for this kind of object (or issue, person, cause, project, etc.), and (b) to decide on this basis to foster these motives, in oneself. Caring thus includes a kind of self-monitoring of one’s ‘outward’ motivational
character that is guided with reference to what is cared about. Moreover, in Chapter III we also saw that care in this sense exhibits what I called the general \emph{projective} structure of motivation, in contrast to the teleological structure exhibited by "desires" of all substantively definable kinds. Together, these two findings point towards an analysis of identification in terms of "projections," or states in which the agent \emph{makes} something her end by giving it to herself as a goal to be pursued, instead of it becoming her end through her cognitive apprehension of its value for her well-being or power to satisfy subjective preferences and needs she already has.

In general, however, the sphere of projective motivation is much \emph{wider} than identification. Projection can be haphazard and even tentative. Neither projecting some end for ourselves, nor the considerations which rationalize or ground our projecting it, automatically \emph{identify} us with our resulting projective motivation to pursue this end in the strong substantive sense of meaning that we \emph{care} about this end, or that caring about it has become an authentic part of our inward character. Rather, in accordance with Frankfurt’s suggestion, by projecting an end, we may acquire a first-order motive to pursue this end that we cannot or do not in the end affirm as authoritative in this deeper sense.

Consider the following type of case. Alfred forms the project of dating Cheryl seriously, not simply because he is physically attracted to her, nor just because he judges that her personality would complement his and add to their \emph{eudaimonia}, but for other reasons as well. These could be ethical reasons, such as that Cheryl lost her husband a year ago and Alfred feels someone should try to fill that role for her and her child, or ‘personal’ reasons, such as a vague sense that it is time for him to start dating more seriously, or still other reasons relating to their mutual past, etc. In other words, Alfred’s motivations to pursue her can at least \emph{include} projective motives. Let us imagine that Cheryl is receptive and things
are working out: as a result of projectively motivating himself, Alfred gets to know Cheryl better, likes her company, is happy in her presence, cares for her well-being, becomes more attracted to her sexually, etc.: in short, he is on the way to ‘loving’ her in a conjugal sense. But then he discovers that Cheryl is his long-lost cousin, or something equally disastrous: though the motivation of his project remains, along with the ancillary desires and emotions which its partial realization has helped foster, Alfred now explicitly does not identify with this project: it becomes ‘alien’ to him.

Yet in spite of this kind of possibility, it does seem that Alfred was at least tacitly identified with his project up until the point of forming explicit volitions against it. The change he undergoes is somewhat analogous to that experienced by the person whose friend points out his amorous ulterior motives for deciding to visit his dentist (see §IV.3.1 above). In general, it seems likely that projective motivation of any sort at least carries with it our tacit identification with the project formed, and may usually develop into positive identification unless its pursuit or realization turns out to be incompatible other things we already care about (e.g. Alfred strongly identifies with his aversion to incest, or cares about adhering to the social mores forbidding it). Because they are actively formed in a way unlike any other kind of first-order motive, projections are qualitatively distinct enough to carry with them a natural affinity to attract the agent’s identification; motivational states of this kind tends to provide focal points around which cares grow up.

But the positive identification involved in full-fledged caring requires a special kind of

---

887 Suppose he had this aversion but did not already identify with it. Then the quasi-Oedipal revelation of Cheryl’s identity would only precipitate in him a conflict between this aversion and his projective motivation to keep pursuing his relation with Cheryl. He could then remain wanton towards these conflicting motives, or strongly evaluate one of them as superior and decide to side with it on this basis.

888 Suppose Alfred’s project has already led to his caring about Cheryl in this further sense, with its more-than-tacit self-commitment. Then, granted that he cares about avoiding incest, his discovery may prompt a real conflict of cares, making him volitionally "ambiguous" until it is resolved.
projection. Suppose that, without ambiguity, Alfred ceases to identify even tacitly with his project of pursuing Cheryl once he discovers their blood relation. Alienating this existing projective motive, must involve forming the project of breaking himself from projecting this end, as well as breaking himself from most of the emotions and desires which this project helped foster. To take the inverse case, suppose he decides that the prohibition on cousins marrying is outdated and does not apply to them anyway, since he and Cheryl do not intend to have children. Spurred by this crises to examine his motives, Alfred’s evaluation and decision to side explicitly with his project helps solidify his care for Cheryl and their relationship, identifying him with his existing motivation to pursue their romance.

Resolutely caring about remaining with Cheryl in this way is distinguished from Alfred’s first project of pursuing her because it involves in addition the higher-order project of continuing to project this end, or of deepening and strengthening the motivation that he had already actively formed.

On this model, the identification involved in caring is constituted by projective motivational attitudes whose ends are themselves motivational states, including first-order projects (or S4 states). Roughly and provisionally, identification is constituted by the satisfaction of the following two conditions:

1. the agent projects the end of developing, maintaining, or strengthening a given motivation that is either already actually hers or is psychologically possible for her, and

2. the agent’s ground for projecting this internal end includes a persisting affirmative strong evaluation of the motive and its associated reasons (a strong evaluation that can itself have various bases, as indicated at the end of Chapter III).

This account fulfills the requirement of qualitatively distinguishing identification from mere second-iteration desires in the subjective (S2) sense of brute wants or preferences, since it proposes that the motivational state constituting identification is one structurally different in
kind from S2 desires. Second-order projection is distinct from second-order desire. It also meets what I called the ‘Qualitative Condition,’ since the project_2 of pursuing a motive_1 (a) must have a kind of evaluative ground that would also dispose the agent to affirm the project_2 itself if it were challenged, and (b) this project_2 will be naturally *self-reinforcing*, in the sense that pursuing this project_2 will help perpetuate and make dispositional the very evaluation of motives_1 that grounds it: the more we work at projecting_1 our end, the more committed we are to ruling out alternative ends, and hence the less revisable our evaluation of the project_1 becomes. As a result, the agent is ‘always already’ identified_2 with this kind of a project_2 to establish actual or potential projects_1 as stable parts of his or her outward character. There is no need to postulate yet higher-iteration projects_3 to affirm our projections_2 concerning motives_1, since these projections_2 are inalienable (though again, they may not be wholehearted).

This model helps capture the reflexive or intrapersonal aspect of caring that Frankfurt has noticed, and explain why caring is so closely related to identification_2. In addition, by relating identification_2 to the kind of motivation through which we care about our own character_1, it avoids the problem of instantaneous or sporadic identification_2 (Obj.B4). Since a *decision* to project an end may not be carried out, deciding to form a project_2 and being projectively motivated_2 are not the same thing. As a result, while this account does analyze identification_2 as a kind of motivation in which the agent *takes initiative* with regard to her motives_1, it does not rest identification_2 in any bare act of deciding. Without being ‘voluntaristic,’ it helps explain how the agent is inherently *active* in her identifications_2. The
active/passive distinction traces in part to the different ways in which persons can be motivated to pursue ends: motives with a projective structure are by nature more ‘actively self-determined’ than other kinds of intentional tendencies towards ends, and this gives us an objective basis for the phenomenological experience of a difference between the ‘internality’ and ‘externality’ of our first-order states.

The motivation which results from projecting ends is something enacted, and this is precisely why, as argued in Chapters I and II, ‘projective motivation’ constitutes the will in its distinctive sense. In forming reflexive projects, or projecting ends concerning her own motives, the person is thus forming distinctively volitional attitudes towards her outward character. According to the projective analysis, then, identification consists in a special state of the will in its thick sense as a kind of motivation for decisions about actions:

specifically, we identify ourselves through higher-order projections. This kind of self-commitment thus truly deserves the description "volitional identification." This analysis remains within the family of hierarchical theories, since it requires states that are higher-order in the weak or thin sense, but it grounds their qualitative authority in the distinctive motivational structure characteristic of "volitions" in the proper sense.

Thus the projective analysis of care seems to me to remedy all the ways in which Frankfurt’s initial model was deficient in explaining our experience of identification as an authoritative intrapersonal relation: it resolves the major phenomenological objections B2-B5. But I have not yet made clear whether this projective analysis can provide adequate conditions for moral responsibility, and thus answer the rather different concerns of Obj.B1. For Fischer could ask whether my view leads to another "mesh theory" of responsibility, subject to a version of the same general objection: could not our projections of motives be caused by responsibility-undermining mechanisms, such as brainwashing or electronic
My response to this concern is threefold. First, this theory is not ahistorical, since identification-constituting projections must have the dispositional form of a persisting inward character: in short, identifications are cares. Second, unlike mere second-iteration desires, it is essential to projections of any order that they are actively formed, or that the agent can regard them as grounded in her own decision to project ends for herself, or regard them as having an intentional explanation in the considerations and dispositions on which her end-forming or self-motivating movement was based. This means that it is built into projections that they are agent-responsive and reasons-responsive. Thus the agent always has what Fischer calls "guidance control" over her projections: if this is compatible with their also having a deterministic causal explanation that traces to laws and events outside her control, then an account of moral responsibility in terms of identification projectively constituted would be in a similar position to Fischer's own account. But third and finally, it could turn out that for them to be agent-responsive in the requisite way, the person must also have what Fischer calls regulative control over her identifications, or be able to 'identify stimulation'?890

890 For Fischer claims that Frankfurt's refined model of identification as decisive judgment-based volition is still subject to the objection that the "second-order volition and the associated judgment could also be produced in a responsibility-undermining way," and that this theory remains "purely structural...and ahistorical" (Fischer, The Metaphysics of Free Will, p.208). But it is clear that Fischer is not taking into account here that Frankfurt's later work points towards a care-structure for identification which does have a temporal dimension; although this model does not make moral responsibility for actions depend on their causal history, it does relate it to the agent's historically sustained character. Furthermore, if cares were directly induced, they would not be reasons-responsive; and to the extent that they depend on strong evaluations which could apparently have external causes, Fischer's actual-sequences theory responsibility has the same problem, as he admits (p.209).

891 Ibid, p.132. However, identifications may have a somewhat different kind of reasons-responsiveness than the "weak reasons-responsiveness" which Fischer singles out as key for moral responsibility (p.166-7). For Eleonore Stump and others have given reasons to think this condition is too weak. However, spelling out the exact way in which different kinds of projections may be reasons-responsive is an elaboration that does not have to be completed at this initial stage of the projective analysis of identification.
otherwise’ than her actual cares dictate. If so, it would mean that identifications and cares involve libertarian choices that by definition could not be programmed or directly caused by external factors. In other words, as I eventually hope to argue, we would have to accept a libertarian resolution to Obj.B1.

IV.6.3. Conclusion: Heidegger and Authentic ‘Care of Self’

If the foregoing projective account of identification, proves to be sustainable, it provides a new basis for Frankfurt’s original thesis that it is the structure of the will that is distinctive of persons. For according to my account, it is primarily through our will that we form our identity, and our deepest selves consist precisely in our intrasubjective volitional activity. Persons are essentially projective beings, who not only give themselves ends for action, but who also project goals about the pattern of their own first-order motives, including outwardly directed projects; persons thus necessarily take some stand with respect to possibilities of their own outward characters (even if only by fleeing the responsibility for self which this capacity implies). Equivalently, caring, which always involves projections that at least obliquely concern ourselves, is the activity which is essentially characteristic of persons as such.

In conclusion, it is worth briefly comparing these results to Heidegger’s determinations about "Dasein" (roughly, entities with individual human existence) in Being and Time. In his discussions of authenticity, Existenz, projection, and care, it becomes apparent that Heidegger understands personhood as constituted by a reflexive structure as Frankfurt

\[\text{\footnotesize (continued...)}\]
claimed later. Moreover, there are reasons to regard this as a *volitional* relation to oneself, although Heidegger avoids using this terminology.

Heidegger introduces the concepts of "authenticity" and "inauthenticity" at the beginning of Division I, following the Second Introduction to *Being and Time*. He begins the first section with the famous hypothesis that "The ‘essence’ of Dasein lies in its existence" [Existenz], because it is not a substance or an entity that simply is "what it is." As I explained in the Introduction, substantial entities merely "have" a possibility as their "property," whereas Dasein’s possibilities are an issue for it. Thus Dasein is not a substance in the Aristotleian sense, and the only ‘whatness’ or "Being as-it-is [So-sein] which this entity possesses" is a being without substantial determination —hence ‘Existenz.’ Like *nous* in Aristotle’s *De Anima*, which acts as a "mirror" of everything rather than having its own determinate nature, Dasein "discloses" all beings, including its own being, which means that its Being involves an indeterminacy unlike that of substances.

This also explains Dasein’s special kind of individuality. Whereas entities defined by their substantive ‘what’ are simply *instances* of a species-form (individuated either by matter, position in spacetime, haecceity, or whatever), Dasein is always mine because it is individuated by its own unique relation to its potentialities-for-Being. It gains this existential individuality from experiencing possible ways it could be in life not just as possibilities in the neutral sense (i.e. as objective qualifications of some entity, like the

---

893(...continued)
Scheler in mind) in his "Letter on Humanism."
894Heidegger, *Being and Time*, Division I, Chap.1, §9, H42, p.67.
895Ibid, H42, p.68.
896Ibid, Division I, Chap.1, §9, H42, p.67.
897Ibid, H14, p.34.
stone), but as possibilities that it owns, possibilities for it, whose limits constrain it. In this relation to its own modal reality, possibilities gain the existential significance of possible choices, achieved actual states of being, and external (non-volitional) necessities gain the existential significance of facticity.

The fact that Dasein is not simply some species of things or a "special case of some genus of entities present-at-hand" thus consists in the fact that its Being matters to it, or has meaning in terms of an undetermined future. As Sartre would later put it, human consciousness stands at a kind of inward distance from its Being (unlike "Being-in-itself"); for Heidegger, its Being includes this difference between Dasein and its Being. This difference arises again from Dasein’s unique relation to possibility:

...in each case, Dasein is mine to be in some one way or another. Dasein has always made some sort of decision as to the way in which it is in each case mine. That entity which in its Being has this very Being as an issue, comports itself towards its Being as its ownmost possibility. In each case, Dasein is its possibility, and it ‘has’ this possibility, but not just as a property, as something present-at-hand would.

Unlike a stone, for example, which has possibilities without them meaning anything to it, Dasein stands in the meaning of its possibilities: it ‘is’ its possibilities in the sense that it lives them as chosen, rejected, not yet determined, or not even faced, and so on. There is a reciprocal interrelation here: Dasein’s possibilities include possibilities of its comportment to other things and to itself, yet it has taken up some comportment to these possibilities, and thus apprehends these very possibilities in light of the comportment in which it already lives.

This special relation to modality thus lies at the heart of the disclosive relation to Being which separates Dasein from Aristotleian substances. The possibilities (and necessities constituted by the limits of these possibilities) which it lives have practical meanings for it,

---

898Ibid, H42, p.68.
however vague or unthematized they may be:

In determining itself as an entity, Dasein always does so in light of a possibility which it is itself and which, in its very Being, it somehow understands. This is the formal meaning of Dasein’s existential constitution.\footnote{Ibid, H43, p.69.}

Accordingly, the "understanding" in which we disclose the "world" —the encompassing relations of significance within which things present-at-hand, equipment ready-to-hand, and other persons acquire their meaning— is affected by our own orientation or "whole basic state of Being-in-the-world," and we always apprehend this whole of this world we are "in" according to the possibilities revealed in it by our basic orientation.\footnote{Ibid, Chap.5, ¶31, H144-145, p.184.} This is due to the fact, Heidegger argues, that

the understanding has in itself the existential structure which we call "projection" [Entwurf]. With equal primordiality the understanding projects Dasein’s Being both upon its "for-the-sake-of-which" [i.e. itself as a determiner of its meaning and its ends] and upon significance, as the worldhood of its current world....Dasein is thrown into the kind of Being which we call "projecting." Projecting has nothing to do with comporting oneself towards a plan that has been thought out, and in accordance with which Dasein arranges its Being. On the contrary, any Dasein has, as Dasein, already projected itself; and as long as it is, it is projecting. As long as it is, Dasein always has understood itself and always will understand itself in terms of possibilities.\footnote{Ibid, H145, p.185.}

As I read this, the way our world appears to us, including the possibilities that seem open to us within it, is ‘colored’ from the outset by an orientation towards ends which we have already assumed, without necessarily thematizing it or being reflectively aware of the commitments (or lack thereof) that determine this orientation. This whole orientation Heidegger calls our existential "mood" or "state-of-mind:"

As existentialia, states-of-mind and understanding characterize the primordial disclosedness of Being-in-the-world. By way of having a mood, Dasein ‘sees’

\footnote{Ibid, H43, p.69.}
\footnote{Ibid, Chap.5, ¶31, H144-145, p.184.}
\footnote{Ibid, H145, p.185.}

386
possibilities, in terms of which it is. In the projective disclosure of such possibilities, it already has a mood in every case.\textsuperscript{902}

Heidegger later expresses this condition by saying that Dasein is "thrown projection,"\textsuperscript{903} or always limited in its possibilities for action by the ways it has already projected itself. However, its projective nature also gives Dasein the possibility of understanding itself "in terms of its ownmost potentiality-for-Being," rather than in terms of the world whose significance and possibilities are already tacitly conditioned by its existing projections.\textsuperscript{904} In other words, Dasein can become authentic by explicitly considering the overall direction or mood of its existing projections, and how they shape its world, and consider alternatives at this inward level: it can positively address its potentiality-for-being (its projective nature) by forming projects concerning its own projections, or (in Heidegger’s language) actively project itself onto certain of its ownmost possibilities for Being-in-the-world rather than others. In this way, it authentically guides itself by what it actively takes responsibility for caring about.

Of course, Heidegger does not use "projection" [\textit{Entwurf}] in a volitional sense as self-motivation, as I have: "Dasein’s Being reveals itself as care" [\textit{Sorge}], in a sense that is prior to "wish, will, addiction, and urge," which are "founded on it."\textsuperscript{905} Projective motivation in my explicitly active sense is thus rooted for Heidegger in an entire futural orientation that is neither completely passive nor optionally chosen: care in this existential sense is Dasein’s

\textsuperscript{902}Ibid, H148, p.188.
\textsuperscript{903}Since \textit{Entwurf} already has the sense of ‘throwing’ something in German, this expression can be understood as something like ‘projected projection.’
\textsuperscript{904}Ibid, Division I, Chap.6, ¶44, H221, p.264.
\textsuperscript{905}Ibid, ¶39, H182, p.227.
"structural whole," as becomes clear in the phenomenon of anxiety. Dasein’s being is always "an issue" for it, because its understanding of the world and (in conjunction) its sense of itself rests on "self-projective Being towards its ownmost potentiality-for-Being," because it already (or "factically") so-to-speak ‘occupies’ some of the possibilities that it has essentially, or has tacitly comported itself towards its general potentiality, "Dasein is always ‘beyond itself’" or runs "ahead-of-itself" in being in a world of things and other persons construed a particular way. This structure, which Heidegger calls "care," thus embraces all essential aspects of Dasein’s being, including its being invested in a entire web of relations of significance involving items ready-to-hand and others (rather than being an "isolated subject"). For example, "willing," in the sense of making something our end, requires this whole background structure of ontological care, since it depends on a world in which ‘objects’ of will and their possibilities are understood, and on the general "for-the-sake-of-which[ness]" or "Being-ahead-of-itself" of Dasein, whereby Dasein can project itself towards "a possibility of the entity ‘willed’.

"Care" as this basic ontological condition of Dasein’s ‘towardness’ or, as we might put it, projectivity, thus does not "stand primarily and exclusively for an isolated attitude of the ‘I’ towards itself" (and if care meant self-concern in this sense, he points out, "the expression ‘care for oneself’...would be a tautology"). But an ability to concern ourselves about our own or projective nature or existential care-structure is open to Dasein: this one particular sort of proactive initiative, itself grounded in care-in-general, is necessary for Dasein to...
become "free for authentic existentiell possibilities." My care can either be "absorbed in the everyday multiplicity and the rapid succession of that with which one is concerned," in which case I forget and flee my Self, falling into the world and apprehending myself only as something "constantly selfsame but indefinite and empty," like the Kantian I-think, or caring can turn inwards towards itself, giving constancy to its concerns through what could be called an ‘intra-projective’ relation:

Selfhood is to be discerned existentially only in one’s authentic potentiality-for-Being-one’s-Self — that is to say, in the authenticity of Dasein’s Being as care. In terms of care, the constancy of the Self, as the supposed persistence of the subjectum, gets clarified. But the phenomenon of this authentic potentiality-for-Being also opens our eyes for the constancy of the Self in the sense of its having achieved some sort of position. The constancy of the Self, in the double sense of steadiness and steadfastness, is the authentic counter-possibility to the non-Self-constancy which is characteristic of irresolute falling.

As I read this, authentic caring is a reflexive mode of care as such; through resolute concern about our projects, we gain an inward "Self" and give it constancy not only in the minimal sense required for any identity over time, but in the thick sense of a lasting volitional character. Existential identity, then, is constituted by caring about our being-as-care, or projecting ends concerning our very projects, emotions, desires, and moods — just as in the model of volitional identification I presented above. The possibility of care in this authentic self-constituting sense is founded (like the inauthenticity which is Dasein’s default mode) in the basic existential care that is Dasein. Moreover, inauthentic Dasein performs only "tranquilized willing," because it passively accepts whatever conventional limits are

911 Ibid.
912 Ibid, Division II, Chap.3, ¶64, H321, p.368.
914 Ibid, H323, p.370: "Care does not need to be founded in a Self. But existentiality, as constitutive of care, provides the ontological constitution of Dasein’s Self-constancy, to which there belongs, in accordance with the full structural content of care, its Being-fallen factically into non-Self-constancy" (italics omitted).
given in everyday public culture, and so "no positive new possibilities are willed." By contrast, in authentically turning our existential care or projective capacity towards itself, we discover possibilities of ourselves that were blocked when we lived in 'wanton' inauthenticity: i.e. the kinds of possibilities which I described as making up the inward or higher "personal world" (see §6.1.B). Heidegger’s sees these two aspects as already implicit in the classical "Myth of Care." As Heidegger remarks, in his last epistle, Seneca says that the good of man is "fulfilled by care (cura)," which has a double-meaning:

Man’s perfectio—his transformation into that which he can be in Being-free for his ownmost possibilities (projection)—is ‘accomplished’ by ‘care.’ But with equal primordiality ‘care’ determines what is basically specific to this entity, according to which it has been surrendered to the world of its concern (thrownness)... Heidegger’s conclusion that ‘care’ in a non-ontic sense (i.e. distinct from mere ‘worry’) is the distinctive ontological feature of Dasein or personal being thus seems to have a clear Stoic lineage: it amounts to something like the idea that ‘Man’ is the volitional entity (in the thick sense of will that I have analyzed in projective terms).

This double-potentiality of ontological care is part of what makes Dasein individual. As Heidegger says at the outset, the special projective character of Dasein’s relation to modality in general gives it two overarching ways of relating itself to possibilities:

And because Dasein is in each case essentially its own possibility, it can, in its very Being, ‘choose’ itself and win itself; it can also lose itself and never win itself; or only ‘seem’ to do so....As modes of being, authenticity and inauthenticity....are both grounded in the fact that any Dasein whatsoever is characterized by mineness.

Heidegger’s analysis differs from Frankfurt’s in this important respect, then: Heidegger sees

---

915 Ibid, Division I, Chap.6, ¶41, H195, p.239.
918 Ibid, H42, p.68.
that the two possibilities of (a) remaining wanton in Frankfurt’s sense or (b) authentically facing (in higher-order care) the question of what to care about, are both built into the ultimate structure of human existence as its cardinal possibilities. The outward focus of a wanton who inauthentically lives in the flux of first-order motivations and concerns (and tacitly identifies with doing so) is also a mode of personhood, revealing the same underlying existential structure that authenticity also depends on.\footnote{919} 

Overall, however, this brief analysis suggests a convergence between Frankfurt and Heidegger on the experience of intrapersonal identification: care in Frankfurt’s sense is roughly similar to "authentic care" or "care for self" in Heidegger’s account. Precisely what Heidegger’s theses about projection and care mean, of course, remains a controversial question to which much scholarly literature has been devoted, and I have not argued that my earlier proposals for relating these concepts to volitional identification fully capture Heidegger’s intention. But I can content myself here with the hypothesis that this analysis prepares the ground on which Heidegger’s theses can be interpreted in a new light, with the possibility of a better understanding of their inner significance and implications than was available to readers working without any prior attention to the phenomena of volitional identification. This background may provide an indispensable framework for grasping the intrasubjective relations involved in Heidegger’s fundamental ontology of Dasein.

\footnote{919}Only Dasein can be "inauthentic." An animal that is incapable of care of the self is not wanton, since it is not fleeing from its capacity for identification; it is not comported towards these options as potentialities of its Being, since it does not exist as “care” in Heidegger’s sense.
CHAPTER V

A Critique of Several Alternative Accounts of Volitional Identification
Contents of Chapter V

V.1. Introduction: Categorizing the Alternatives

V.2. Identification as Rational Evaluative Judgment
   2.1. Eleonore Stump: Identification as Thomistic S3-Desires for Motives
   2.2. Charles Taylor: Identification as Strong Evaluation
   2.3. Gary Watson: Identification as First-Order Valuation
   2.4. David Velleman: Identification as rooted in the Desire to Act Rationally

V.3. Identification as Self-Discovery: Bransen and Tugendhat

V.4. Buss’s Nonattitudinal Account of Identification via Practical Reasoning

V.5. A Critique of Dennett’s Neo-Behavioralist Account of Identification
   V.5.1. Introduction: Dennett’s Cognitive Approach
   V.5.2. The Intentional Stance: Pragmatism, Neo-Behaviorism, Reductionism
   V.5.3. Dennett’s Reflective Self-Manipulator Model of Identification

V.6. Conclusion: Volitional Identification and Personhood
   V.6.1. The ‘Mentalist’ Tradition Behind’s Dennett’s Pragmatist Interpretation
   V.6.2. Why the Radical Implications of Frankfurt’s Analysis Were Missed
V.1. Introduction: Categorizing the Alternatives

As Frankfurt once remarked, "The problem of explaining identification is not, of course, peculiar to the hierarchical model," since it "must be dealt with by any account of the structure of volition." The phenomena of intrasubjective identification, alienation, ambiguity, wholeheartedness, and so on, are real and ineliminable, and whether or not they correspond exactly to any concepts in traditional ‘folk psychologies,’ our experience of them is prior to theoretical attempts to explain these phenomena. As I argued (in IV.3.1), the central significance of these phenomena for our understanding of personhood, character, and the capacity of agents to shape their ‘selves’ cannot be dismissed.

The previous chapter traced the development of Frankfurt’s theory of identification through his analysis of care, and argued that identification per se (or identification₁) should not be analyzed in terms of the distinct concept of decisive or wholehearted identification. I also argued that the identification₁ involved in care could best be understood in terms of a reflexive kind of projective motivation, in which the agent takes actual or possible states of her own outward character and makes them ends for herself (as opposed to desiring them either because of their subjective preferability or objective value for her well-being). Since this account fits some aspects of Heidegger’s analysis of Dasein, I called it the existential account of identification₂, which understands care in terms of states that are distinctively volitional (in the original projective sense of willing, as argued in Chapter II).

But, largely in response to the range of objections against Frankfurt’s initial model of identification₂ that I described in the previous chapter (IV §3.2), several alternative accounts of identification have been proposed since the hierarchical model appeared. In recognition of the regress and ab initio problems especially, virtually all of these accounts understand the

---

920 Harry Frankfurt, "Identification and Wholeheartedness," reprinted in The Importance of What We Care About, p.166, note 9.
conditions of identification as *qualitatively* distinct from desire in the sense of preferences, urges, or inclinations.⁹²¹ In accordance with classical anthropologies, Frankfurt’s interlocutors have most often proposed that identification is constituted by some form of *rationality* in one’s motives, or some kind of practical judgment that is supposedly more *authoritative* for the agent than mere desires: thus they have opposed the nascent emotivism of Frankfurt first model with various rationalist alternatives. In different ways, Eleonore Stump, Charles Taylor, Gary Watson, Susan Wolf, and David Velleman have all adopted this kind of strategy.

Several other approaches are also possible, however. Sarah Buss presents an account of autonomy that dispenses with idea that any higher-order attitude is required to make the intentions on which an agent acts *her own*; she argues instead that whether an intention is autonomous depends on its relation both to the agent’s good and to her character. This account could also be classed with the rationalist accounts, except for its greater emphasis on the agent’s existing character. Alternatively, following Ernst Tugendhat, Jan Bransen has proposed to interpret identification in terms of a process of *self-discovery*. There are also positions that would make identification depend on rich *interpersonal* conditions of evaluation and attribution: Joel Anderson and Cheshire Calhoun (in her work on integrity) argue for different versions of this approach.⁹²² Finally, Daniel Dennett has proposed a pragmatist analysis in which identification depends on objective attribution, but *within* the individual: the agent takes a type of third-personal attitude or diagnostic attitude towards himself.

---

⁹²¹ As we will see, though, Dennett’s account fails to satisfy the Qualitative Condition.

It is also possible to further subdivide these alternatives according to whether their version of identification still requires states or attitudes that are higher-order in the weak or thin sense of being about first-order desires, passions, emotions, intentions, and so on. Taylor, Wolf, Stump, Dennett, and possibly Calhoun maintain hierarchical reflexivity as one element of their account, whereas Watson, Vellemen, Buss, and Bransen, do not. Although this is not the most salient way of classifying alternative approaches, since it obviously lumps together radically different views, the issues it raises will come up in my critiques of the views that dispense with reflexivity or thin hierarchicalism altogether.

It is essential to the project of a reconstructed existentialist account of personhood to resist not only emotivist, but also these rationalist, interpersonal, and pragmatist alternative accounts of volitional identification. For only when the weaknesses in all these explanations are shown can an existentialist appeal unambiguously to the phenomena of identification as the decisive evidence for her distinctively volitional account of personal existence. Accordingly, in this chapter, I will argue that there are fatal flaws with every one of these alternative proposals, except perhaps with Bransen’s (which, when properly interpreted, may not be inconsistent with the existentialist approach I outlined in Chapter IV).

IV.2. Identification as Rational Evaluative Judgment

IV.2.1. Stump’s Thomistic Analysis of Identification as S3-Desires for Motives
In her paper on Frankfurt’s concept of free will, Eleonore Stump argues that to avoid the familiar regress problem [Obj.B2] with Frankfurt’s original hierarchical account of identification, we must account for why "an agent cannot be a passive bystander to his second-order volitions," or why he is identified with them without needing them to be approved by yet-higher order desires. She believes that the best prospect for satisfying this Qualitative Condition (IV §3.2), is to reconceive higher-order volitions as rational evaluations of the good:

In the spirit of Aquinas’s view of intellect’s direction of the will, we can make the following first revision of Frankfurt’s account of the self. An agent has a second-order volition V2 to bring about some first-order volition V1 in himself only if the agent’s intellect at the time of the willing represents V1, under some description, as the good to be pursued.

Stump emphasizes that this is not meant to be a normative condition of rationality: a volition V2 is formed by the agent’s "intellect" in Aquinas’s sense as long as there as "some chain of reasoning" representing it as the good "would figure in the agent’s own [honest] explanation of his action." This condition can be met even if the agent’s reasoning was "hasty, thoughtless, ill-informed, invalid, or in any other way irrational." This account simply requires higher-order volitions to meet the teleological condition of representing the states to which they convey identification as states that are good in some sense.

However, given her Thomistic intentions, Stump presumably means by this that the agent must by able to conceive V1 (and thus V2 as well) as good for the agent in an objective eudaimonistic sense (i.e. not simply as subjectively preferred). In other words, for a volition

---

925Ibid, p.216.
926Ibid, p.216. Stump also points out that in this sense, the agent can have reasons for his higher-order volition even though "he has not consciously formulated" them, as long as these are the reasons he would recognize if asked.
V2 to have an *intellect-source* in Stump’s fairly non-normative sense is for it to be a *desire for V1* in something like my S3-sense of desire. According to Stump, for a motivational state to have an intellect-source in judgments about its objective value for well-being is sufficient to make it *inalienable*:

On the revised [Thomistic] account, an agent forms second-order desires by reasoning (rationally or otherwise, consciously or not) about his first-order desires; and a second-order desire is a direct result of an agent’s intellect representing a certain first-order desire as the good to be pursued. Given this connection between intellect and second-order desires, an agent cannot be a passive bystander to his second-order volitions.

Stump develops this picture somewhat in a more recent paper, using the example of Francesca, who identifies with her desire to stand at a faculty meeting on behalf of censuring a colleague she believes is guilty of sexual harassment, and acts on this desire. On the Thomistic account,

the first-order volition to stand reflects the determination of Francesca’s intellect that in these circumstances and at this time the good thing to do, in some sense of ‘good,’ is standing. Her second-order volition also stems from a determination of intellect. In a reflective review of states of her will, her intellect determines that the first-order volition to stand is good. In making this determination the intellect is not considering just what it would be good to do on this occasion and in these circumstances....Rather, in reflecting about first-order volitions, Francesca’s intellect is reflecting on what sorts of volitions it is good for Francesca to have [in general].

The second-order volition reflects an "all-things-considered judgment of her intellect in a way that her first-order volition does not," since the latter can be hasty and event-specific, whereas the former is based on a more thoroughly considered judgment about "what would

---

927 As she says, according to Aquinas, "The will is a natural inclination or appetite for the good; and the intellect moves the will, without coercion, by showing it what the good to be pursued is in a particular set of circumstances," though the will (and the person’s acquired habits) also indirectly move the intellect to consider some things more than others in making its judgment (p.216, note 8). Though Stump does not flesh out "the good to be pursued," since Aquinas understands it as the condition for *flourishing qua human*, I will assume that Stump has in mind a judgment like that described in my schema for S3-desires.


be good to do in general in circumstances like these." Stump believes that defining second-order volitions as *qualitatively distinct* from other states in this way (though still second-order) explains why we are automatically identified with second-order volitions so analyzed:

A person identifies herself with her second-order volitions, then, because they reflect the whole view of her mind in a way that first-order desires do not, and because (in a different sense of identification) she is to be identified not only with her will but also with her mind....

An agent’s second-order volitions are authoritative for her because they reflect the all-things-considered judgment of her own mind, and her mind is constitutive of her.

In short, the agent who identifies *is* the connection between this particular mind and will as defined on Aquinas’s theory of motivation, and deriving from judgments that reflect the whole state of this infinitely authoritative (because agent-constitutive) source is what makes an evaluation of motives, itself inalienable and identification-conveying.

Quite a bit rides for my purposes on whether such an intellectualist analysis of identification is convincing. The basic constrast I set up in Chapter I between a Thomist account of will as ‘rational appetite’ for the agent’s good, and an existential account of will as projective motivation, is *tested* in the crucial case of volitional identification by the question of whether a Thomist or projective account of identification (as sketched in IV §6.2) is more phenomenologically adequate. This test case is particularly helpful for the existentialist cause, because it is abundantly clear that identification *cannot* be constituted as it would have to be on a Thomistic approach. To show this, I will lay out three problems for

---

930 Ibid, p.202. We might add that for Stump, the second-order volition seems to require a review of the different *senses* in which an act or motive might be good, and some weighting of them, while different first-order motives can derive from an intellectual focus on a *single* sense of ’good’ at the moment.

931 Ibid, p.202 and p.203. Thus Stump says that conflict among second-order volitions can arise only for agents who are “ambivalent between whole worldviews” (p.203).
Stump’s account.

First, as Stump notes, a Thomistic reinterpretation implies that whether or not an agent represents as the good to be pursued some desire for an action $p$ or state of affairs $q$ "in turn will depend on whether the agent considers $p$ (or $q$), under some description, at that time, a good to be pursued." Stump’s new account modifies this to something like the following: ‘...on whether the agent considers $p$ (or $q$), under some description, as the good to be pursued in this general type of circumstances, taking all types of ‘good’ into consideration.’ If our practical deliberation is eudaimonistic, then, identifications or second-order volitions will only have first-order motivations, as in Pink’s model (see Chap. IV §3.1). But this is incorrect, as we have already seen: it excludes from consideration grounds that might justify the pursuit of an end-state, or warrant motivation towards it, without resting this justification completely in the value of the end-state itself for the agent’s well-being, all things considered. In other words, evaluations of motives, themselves as constituents of character is not sufficiently distinguished from evaluation of first-order actions or action-types on Stump’s account.

Second, this model implies a conception of the self that is on its face question-beggingly rationalistic. As Stump says, her account solves the ab initio problem by saying that an agent is "more truly identified with his second-order than with his first-order desires" because the former are "the expressions of his intellect’s reflection on his will, and the agent is to be identified with his intellect." And with respect to her biblical example, Stump says that

Tamar’s reasoning faculty is essential to her, as no other part of her character is. If she were incapable of emotion, or if she were to become apathetic through

\footnote{Stump, "Sanctification," p.220.} \footnote{Ibid, p.223.}
Thus the ‘real self’ is the agent’s intellect in its capacity for practical evaluation of the eudaimonistic value of motives. But the fact that a capacity for rational deliberation may be one essential condition of personhood (see IV §2.3) does not imply that by itself it is the wellspring of self-constituting identification. Moreover, to the extent that reasoning does play a necessary role in forming identifications, it may not be reasoning that a motive is good in the sense of contributing to overall well-being. Stump’s Thomistic account implies that a person is most essentially their fully considered S3-desires (her constitutive sense of ‘identification’): but this is to presuppose precisely what is at issue in a phenomenology of identification. On the existentialist model, as we saw, projections that carry or convey identification also involve some rational consideration of grounds, but precisely not of the kind the Thomist envisions.

Despite what Stump says, on her account of identification, Thalberg’s ‘Animal Self’ objection (IV §3.1) really does become a serious problem, because the agent’s capacity "to be identified with her darker first-order desires" does not obviously depend on her representing them as good for her: as volitional states like spite illustrate, agents are capable of radically evil identifications that are not motivated by any anticipation of greater well-being. In this sense, the phenomena of identification provide many counterexamples to the Platonic postulate about human motivation which is central to the Thomistic approach. An intellectual judgment that a motive is (in the circumstances) good in the eudaimonistic

---

\(^{934}\)Ibid, p.225.

\(^{935}\)Ibid, p.224. In fact, I suspect that it was precisely a suspicion that Frankfurt’s second-order desires implied Stump’s kind of rationalistic conception of the self that originally motivated Thalberg’s objection. On other models that explicitly avoid equating second-order volitions with intellectual exercises, Thalberg’s objection loses much of its force.
sense is not, then, obviously a necessary condition for identifying with that motive$_1$.

Third and most importantly, an S3-desire for a motive$_1$ is not a sufficient condition for identifying with the motive$_1$ either. Stump argues that

A second-order desire is itself an expression of the agent’s reasoning and therefore *eo ipso* accepted by the agent as approved by his reasoning. If an agent’s reasoning approves a desire as the good to be pursued, it must also in the very same process approve the desire for that desire. An agent who at one and the same time unambiguously considered a certain desire$_1$ as a good to be pursued and also rejected the desire$_2$ for that desire$_1$ would not be sane.$^{936}$

It may be correct, as Stump claims, that an agent cannot sanely evaluate a motive$_1$ as good in this sense without thereby also being disposed to evaluate a desire$_2$ for that motive$_1$ as good for the same reason.$^{937}$ But this is immaterial, because a person can judge both his desire$_1$ and his desire$_2$ for that desire$_1$ as good in the Thomistic sense *without* identifying personally with either of them: I may make this judgment in a purely neutral fashion, as in a third-personal belief that such a motive$_1$ would indeed be objectively good for ‘one’ to act on, or promote human flourishing, without this meaning that *I will to commit myself* to this motive$_1$.

Thus we may add to Frankfurt’s original argument that a wanton may engage in instrumental reasoning,$^{938}$ the crucial point that a wanton may also deliberate on and form practical judgments about her first-order states without ceasing to be wanton, or taking an inward stand with respect to them. Frankfurt has consistently argued that evaluative

---

$^{936}$Ibid, p.222-223.
$^{937}$See my discussion of S3-desires for S3-desires in Chap. II, §3 (9). As I say there, unless something blocks an evaluative attitude from generating the first-order S3-desire it would normally generate, the higher-order S3-desire for that desire would be redundant. This is consistent with Stump’s point that on her model, a third-order desire with an intellect-source could only arise as an independent state if it were “a temporary measure to bring [the agent’s] second-order desire into line with a change in his reasoning” (p.221).
$^{938}$See Frankfurt, ”Freedom of the Will and the Concept of a Person,” p.17.
approval and disapproval do not by themselves constitute identification or alienation.\textsuperscript{939} For we can recognize the normative status or objective authority of even a ‘strong’ value-judgment, and yet regard this evaluation as if from a speculative distance: thus the existentialist will maintain that giving the evaluation first-personal motivating significance, which changes it from a cognition into a volition, requires a further act of deciding to identify with the evaluation or the standards from which it proceeds, i.e. actively committing oneself to the evaluative ordering it implies.\textsuperscript{940}

Moreover, it even seems possible that we could believe acting on a particular desire\textsubscript{1} is objectively the best for us in the circumstances or most conducive to flourishing, yet find that this desire\textsubscript{1} is alien to us, since we actively dissociate ourselves from it by identifying with contrary motives\textsubscript{1}. According to our earlier analysis, this would entail a strong evaluation that did not simply rank our preferred motive\textsubscript{1} over the alien desire\textsubscript{1}, but radically excluded that desire\textsubscript{1} from our internal options (IV §4.2). But such evaluations need not proceed on a eudaimonistic valuation of the two motives\textsubscript{1}; there are other standards (e.g. deontic ones) according to which they may contrast qualitatively or essentially. As Frankfurt recently remarked in what I take to be an insightful statement (and one that clarifies his own position):

Endorsing something is consistent with being too separate and too disparate from it for the endorsement to be tantamount to an identification: identifying with something makes a person one with it, but merely endorsing it does not. Moreover, endorsement implies or at least strongly suggests approval. However, a person may identify himself with aspects of his character of which he actually disapproves...\textsuperscript{941}

\textsuperscript{939}See Frankfurt, “Identification and Externality,” esp. p.65.

\textsuperscript{940}See my paper on “The Meaning of Kierkegaard’s Choice Between the Aesthetic and the Ethical,” §3, esp. p.83.

\textsuperscript{941}From Frankfurt’s argument in response to my recent Eastern APA presentation on "Freedom of the Higher-Order Will" (Atlanta, December 1996), mss. p.3. In several previous papers, Frankfurt has sometimes used the notion of "endorsing" a desire to stand for identifying with it. This remark seems to admit that those (continued...)}
This is important, because Frankfurt is usually read as holding that a person is identified with her desires, by "endorsing" them either in the subjective sense of preferring them or in the objective sense of "approving" them in her evaluative judgments. But, as is clear from Chapter IV, he now recognizes that preferring or approving in these senses is insufficient for identification.\textsuperscript{942} In my terms, it is possible to \textit{project} some first-order state $S$ while disapproving of $S$ in several senses, including believing that $S$ will not contribute to one's overall flourishing. Thus, \textit{pace} Stump, it is false that our "reasoning faculty" determines what we "really want."\textsuperscript{943} A Thomist might object that this would be an \textit{akratic} state. This is debatable, since I might have overriding moral or personal reasons for projecting $S$ despite my low estimation of its eudaimonistic value; yet in any case, the existentialist approach recognizes that akratic cares or dispositions of identification are not only possible, but may represent some of its most familiar forms. Even at the level of personal resolution or commitment, I may devote myself to what I know I should not, or not devote myself to what I know I should.\textsuperscript{944}

In a later paper, Stump defends her analysis with the biblical example of Ruth, who finds it unthinkable to leave Naomi (her mother-in-law), when her husband dies.\textsuperscript{945} Unlike the

\textsuperscript{941}(...)continued

passages were not strictly accurate, and to suggest that we should not read any explanatorily important into those earlier passages.

\textsuperscript{942}It might be suggested that this is because Frankfurt now analyzes identification in terms of "satisfaction." But, interestingly, since being \textit{satisfied} with a desire, entails being satisfied with one's higher-order positive evaluations of it (see IV §5.2), "disapproving" of a desire, seems to be inconsistent with being fully satisfied with it. If so, then the view indicated in this statement ought to lead Frankfurt to \textit{doubt} the satisfaction-analysis.

\textsuperscript{943}See Stump, "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will." p.226.

\textsuperscript{944}The existentialist is in the same position as others in requiring a coherent account of \textit{akrasia} that makes sense of the recognized phenomena while fitting her theory of motivation —an account I have not tried to provide. A phenomenology of identification provides the ground for a full analysis of weakness of will. But I see no reason at this point to presume that the existentialist should have a harder time at this than the Thomist.

\textsuperscript{945}See Stump, "Intellect, Will, and the Principle of Alternative Possibilities," reprinted in \textit{Perspectives on} (continued...)
unwilling addict’s dominant desire, which "is disconnected with his intellect," Ruth’s desire to remain with Naomi is identified with her (even though the desire may be irresistible), because "it flows from her intellect and is dependent on it." But it is not in fact clear, as Stump says, that the unwilling addict’s intellect "at the time of taking the drug doesn’t represent taking the drug, under and description, as the good to be pursued." For in Aquinas’s terms, this would seem to imply that the addict does not even perform a human act when he takes the drugs. But Frankfurt rightly insists that the unwilling addict is acting or guiding his movements by his intentions, not just behaving, in taking the drugs. Contrary to Stump’s suggestion, the addict clearly does represent taking the drug (and thus the desire to take it) as a good to be pursued under some description, even if he also judges this desire unworthy, or opposes it on the basis of a strong negative evaluation of it. Aquinas’s criteria, like Davidson’s, seem inadequate precisely in being unable to accommodate cases where an agent truly acts on her own intention, but nevertheless is not inwardly identified with this intention or action: thus there is an unavoidable irony in trying to explain the phenomenology of identification in Thomistic terms.

Moreover, even if Ruth (in Stump’s example) does judge her desire to stay with Naomi as good in the strongest terms, the existentialist will insist that this cannot be why she is

\footnotesize

945(...continued)
947Ibid, p.249.
949Stump says that the irresistible desire, which moves the addict is different than "desires involved in ordinary incontinence," which move the agent to an act by representing the act as good ("Intellect, Will, and Alternate Possibilities," p.249). Stump has to say this, because if the unwilling addict does act on a desire that represents the drug as a good, then he comes out as incontinent on the Thomistic model; and this is clearly wrong, since an incontinent person is one who is responsible for acting on the motive he does.
identified with it. Consider Frankfurt’s example of Lord Fawn, who cannot go through with his plan to hear a report from Gowran, a man he asked to spy on his fiancée, Lizzie Eustace:

It is Fawn’s judgment that the best thing for him to do would be to speak with Gowran about Lizzie, and he tries to carry out that course of action. At bottom, however, he is unwilling for his will to be shaped in that way.

We could add that Fawn has an S3-desire (with a source in his intellect) to hear the report, but discovers that he cannot identify with acting on this desire. He (apparently irresistibly) finds this desire alien to him, because he cannot will it. This does not mean that his judgment concerning its value has necessarily altered. Rather, that judgment may simply be distanced from him, along with the desire. It is much more plausible to say that this is because he cannot project it as an end for himself. This does not mean that his unwillingness to project it is arbitrary or groundless; rather, it appears to have dispositional grounds that have nothing to do with his perception of what is best according to the measure of his well-being, which is why he can distance from his inner self the judgment underlying the S3-desire in this case.

IV.2.2. Charles Taylor: Identification as Strong Evaluation

In his approach to the issues raised by Frankfurt’s initial hierarchical model, Charles

---

950See Frankfurt, "Rationality and the Unthinkable," in The Importance of What We Care About, p.183. The example comes from Trollope’s novel, The Eustace Diamonds, ch. 59.

951Ibid. Frankfurt adds that someone’s "assertion that the action is for him unthinkable does not require him to maintain that there could be no circumstances in which it would be his best alternative. Nor does it require him to deny that there might be circumstances in which it would be reasonable for him to perform the action....It may be clear to him that there are matters with respect to which he is incapable of acting rationally" (Ibid, p.184, my italics).

952It is somewhat ironic, then, that Stump cites Frankfurt’s "Rationality and the Unthinkable" in connection with her Ruth example ("Intellect, Will, and Alternate Possibilities," p.247, note 18).
Taylor employs his famous distinction between weak and strong evaluation. As we saw, Taylor defines weak evaluations as preferences in the purely ‘quantitative’ sense (allowing for comparisons on a utilitarian scale of ‘more’ and ‘less’ desired), as opposed to strong evaluations in which the alternatives are "characterized contrastively." He points out that "I can have what Frankfurt calls ‘second-order volitions’ on the basis of weak evaluations." This shows the inadequacy of Frankfurt’s initial model: for in weak evaluation, the rejected alternatives are only contingently incompatible with the alternative that wins out as most preferred, whereas in strong evaluation, "the rejected desire is not so rejected because of some mere contingent or circumstantial conflict with another goal...the conflict is deeper." But it is the latter kind of evaluation which is central to personhood on Taylor’s view: thus "what is missing" from Frankfurt’s account is "strong evaluation" of our motivations or "the qualitative worth of different desires."

As we have seen (IV.4.2), in "Identification and Externality," Frankfurt develops his initial account to recognize the role of strong evaluation in qualitatively distinguishing identification from mere preference. Thus, as Frankfurt says later, a utilitarian account on which "the only rational good is well-being" makes every personal commitment corrigible, and thus leaves no room for "moral integrity." As Taylor suggests, this is because the utilitarian tradition implies that our well-being is a function only of preference-satisfaction,

---

953 See Taylor, "Responsibility for Self," reprinted in Free Will, ed. Gary Watson. The main theme of this essay was introduced in Chapter II, §1 and §3.8.
956 Ibid, p.18.
958 Ibid, p.16.
959 Frankfurt, "Rationality and the Unthinkable," p.179.
or that our only honest evaluations are "non-qualitative." Consequently, for utilitarians, human subjects are only simple weighers, whose reflection "terminates in the inarticulate experience that A is more attractive than B" — or who are only motivated by S2-desires, in my terminology. But persons have the reflective capacity to articulate the values of incommensurable options, because they are capable of strong evaluations that invoke a language of qualitative contrasts: "within an experience of reflective choice between incommensurables, strong evaluation is a condition of articulacy." An agent who strongly evaluates her desires thus "characterizes his motivation at greater depth" than a simple weigher. Similarly, caring about our motives, in Frankfurt’s sense involves drawing qualitative distinctions regarding the importance of different ends and principles, and being concerned with "the kind of being we are" in having or acting on these motives, as Taylor puts it.

Thus it also follows from Frankfurt’s modified conception that, in Taylor’s words, "the capacity for strong evaluation in particular is essential to our notion of the human subject." Like Frankfurt, Taylor also thinks that we can thereby become responsible for our character, without our identifications merely resting on an arbitrary fiat, or "radical choice" in Sartre’s sense. He also affirms that our self or identity is "defined by our fundamental evaluations," and includes de facto properties, such as our heritage, only "as assumed in a

---

960 Taylor, "What is Human Agency?", p.23.
963 Ibid, p.25.
965 Ibid, p.28.
966 Ibid, pp.29-32; also see Taylor’s critique of Sartre in his essay, "Responsibility for Self," pp. 119-121, where he argues that radical choice has unwanted emotivist implications.
certain way," or through agent-identification.\textsuperscript{967} This leads Taylor to a conclusion similar to Frankfurt’s notion of volitional necessities that define the most inward self from which a person’s identifications flow: he argues that we are "defined by certain evaluations which are inseperable from ourselves as agents. Shorn of these, we would cease to be ourselves."\textsuperscript{968} The idea is that the capacity of our strong evaluations to define a coherent identity depends on their being restricted by a most fundamental core of evaluations that serve as a kind of basis for our entire sense of self:

The notion of identity refers us to certain evaluations which are essential because they are the indispensible horizon or foundation out of which we reflect and evaluate as persons. To lose this horizon, or not to have found it, is indeed a terrifying experience of disaggregation and loss.\textsuperscript{969}

We might compare this to Frankfurt’s idea that our volitional necessities define our "essence as a volitional creature,"\textsuperscript{970} and that without them we would be utterly anarchic agents, lacking in "the ideals that define the essential nature of a person."\textsuperscript{971} In my terms, then, Taylor’s "fundamental evaluations" would establish a sharp boundary for the range of motivational possibilities that make up the ‘inward personal world’ (see IV §6.1.b-c). It is essential to persons on Taylor’s account that they can entertain some range of alternate possibilities of their own outward character, however:

Whereas for the simple weigher what is at stake is the desirability of different consummations, those defined by his de facto desires, for the strong evaluator reflection also examines the different possible modes of life or modes of being of the agent.\textsuperscript{972}

\footnotesize
\begin{itemize}
\item \textsuperscript{967}Ibid, p.34.
\item \textsuperscript{968}Ibid, p.34.
\item \textsuperscript{969}Ibid, p.35. However, note his discussion of radical re-evaluation of oneself in "Responsibility for Self," pp. 122-125.
\item \textsuperscript{970}Frankfurt, "Rationality and the Unthinkable," p.188.
\item \textsuperscript{971}Frankfurt, "On the Necessity of Ideals," p.25.
\item \textsuperscript{972}Taylor, "Responsibility for Self," p.117.
\end{itemize}
There is thus a large degree of convergence between Taylor’s and Frankfurt’s accounts of intrapersonal identification and decisive identification. But there is one respect—especially important to the existentialist—in which Taylor’s account is inadequate, and in this respect it suffers from a problem similar to Stump’s. Although the notion of strong evaluation is wider than Stump’s Thomist notion of motivation by a judgment of the good to be done (as in S3 desires), and may therefore also include kinds of judgments which typically ground nondesiderative projective motivations, identifications (like projections in general) never follow necessarily from an agent’s having or recognizing such grounds. Thus although strong evaluation may be necessary for identification (as noted in IV.4.2), it cannot be sufficient for it: identification requires a decision about motives, made on the basis of strong evaluation (or, to be completely accurate, an enduring disposition to evaluate motives in a certain way and to decide for or against them on this basis). Thus identification cannot simply consist in a judgment itself (whether based on ethical or aesthetic standards, or other traditional or personal considerations). Judgments of this kind are not essentially inalienable, as the Qualitative Condition requires. The dispositions involved in caring about something (and thus in shaping one’s own will) must involve an established pattern of evaluation-based decisions to make them inalienable. This does not mean that we can establish cares simply by deciding to, but any state of care involves the ongoing distinctively volitional element of decisions about appropriate first-order states.

Taylor’s analysis of personhood is misleading on this point. In his essay, ”The Concept
of a Person," Taylor begins with the idea that a person must be a respondent, i.e. someone capable of replying to verbal meaning. But in contrast to Dennett (as we will see) Taylor understands this in a realist rather than merely pragmatist sense: "animals and human beings are subjects of original purpose. That the cat is stalking the bird is not a derivative, observer-relative fact about it." In other words, human beings and animals really have first-order wills; it is not just that we can attribute such states of will, desire, aversion etc. to them, because things really can "matter" to them, as Taylor says.

As Taylor sees it, the significance which things have for both humans and animals as a result of our "original purpose" with respect to them precedes conscious articulation of this significance: "so original purpose cannot be confused with consciousness." The fact that an animal can respond to the "original significance of things for it" is not a result merely of its "consciousness" of them, as Taylor says. Since, on his view, persons are a subset of "respondents" in this sense, Taylor agrees with Frankfurt that the difference between humans and animals is not essentially a difference in the type of conscious representation of which they are capable. The higher feelings which are essential to human personhood, such as "pride, shame, guilt, sense of worth, love, and so on" (which Frankfurt would call states of care), involve evaluations of situations to which they are a response, and hence different

---

975 Ibid, p.97.
977 Ibid, p.100.
979 Ibid, p.100. Note that Taylor says: "As long as we think of agents as the subjects of strategic action, then we might be inclined to think that the superiority of persons over animals lies in their ability to envision a longer time scale, to understand more complex cause-effect relationships... If we think merely in this strategic dimension, then we will tend to think that this representative power [rational consciousness] is the key to our evolution from animal to man" (p.102). This could serve as an apt diagnosis of Dennett’s motivations for believing that a consciousness-based account can provide the best set of conditions for personhood.
representations of the situation may change the judgments with which these feelings are associated.\textsuperscript{980}

As a result, Taylor argues that being a person in the moral sense, or being capable of these peculiarly human emotions and ends, requires the ability to "recognize and acknowledge," in a "strong" sense, the standards of judgment which these emotions and ends involve. These standards provide a basis for strong evaluations, which involve choices between \textit{essentially} incompatible alternatives.\textsuperscript{981} Unlike animals, who may only follow standards without recognizing them, the ability to recognize of standards of strong evaluation that guide our actions is essential to us as persons:

These human matters are also connected with consciousness in some sense. One could indeed argue that no agent could be sensitive to them who was not capable of formulating them, or at least giving expression to them; and hence the kind of consciousness which language brings is essential to them....To be a moral agent is to be sensitive to certain standards. But ‘sensitive’ here must have a strong sense: not just that one’s behavior follow a certain standard, but also that one in some sense recognize or acknowledge the standard.\textsuperscript{982}

But this formulation has a misleading implication. Things mattering or having significance for moral agents cannot come, as Taylor is suggesting, just from a linguistic \textit{consciousness} of ends that are \textit{natural} to human life, or simply from \textit{understanding} communal standards that form a context for that life. This view is too teleological to take account of the fact, as I argued above, that for any awareness or consciousness of moral standards to give \textit{personal} significance to my practices, I must adopt or \textit{identify} with those standards: otherwise, they can only give meaning and rationality to my actions in a person-neutral sense, or from a third-personal perspective. In Taylor’s terms, I could perfectly well

\textsuperscript{980}Ibid, p.100.

\textsuperscript{981}Taylor, "What is Human Agency," p.21.

\textsuperscript{982}Taylor, "The Concept of a Person," p.102.
see that another person could *interpret* my action as done for the sake of some specially human end (courage, for example), but unless I internalize that reason for the performance as *my own reason*, I have not *acted* courageously in the strict sense.

Thus we must supplement Taylor’s view that "Moral agency...requires some kind of reflexive awareness of the standards one is living by (or failing to live by)." Such a reflexive awareness need not be strategic, but an understanding that standards *apply* to one’s feelings, inclinations, and so on, is not by itself sufficient for them to have the first-person significance of being authoritative for the self or identified with the person’s inner character. Such an awareness or understanding is in the same situation as an reflexive attitude (or mode of affective consciousness) towards one’s own passions. As we saw, Frankfurt holds that an attitude towards a passion can determine its "internality or...externality" only if the attitude itself is one that is essentially identified with the agent. To answer this objection in his own terms, Taylor would have to claim that persons are *by nature* not only conscious of but also identified with ends which moral and communal standards require —a stronger teleological claim than he is willing to make.

This analysis also potentially raises a problem for Frankfurt’s own claims about volitional necessity, however. For persons to be both absolutely passive with respect to evaluations and decisions they cannot but make and be yet *completely* identified with them seems tantamount to saying that the identification which makes us moral agents begins in a metaphysical telos which is simply given in our own individual nature, if not in human nature in general. And this would make it very hard, if not impossible, to see how identification can be the primary measure of moral responsibility for our character. Neither

---

983 Ibid, p.103.
of these options — absolutely arbitrary, instantaneous decision, and absolutely given telos — can convincingly account for the primordial origin of the decisive or highest-order identification characteristic of morally responsible persons. For this reason, the existentialist theory of personhood eventually needs to include its own alternative alternate account of decisive identification or the highest-order will.985

V.2.3. Gary Watson: Identification as First-Order Valuation

In his 1975 paper, after critiquing Frankfurt’s initial account, Gary Watson admits that "There may be something to the notions of acts of identification and of decisive commitment;" as phenomena they are of undeniable significance for the analysis of agency and moral responsibility.986 But he argues that it is

...unclear why these acts of identification cannot be themselves first order—that is, identification with or commitment to courses of action (rather than with or to desires)—in which case, no ascent is necessary, and the notions of higher-order volitions become superfluous or at least secondary.987

This suggestion has some affinity to Pink’s Davidsonian objection A2 (discussed in IV §3.1), since Watson adds that "the same considerations that constitute one’s on-balance reasons for doing some action, a, are reasons for wanting the ‘desire’ to a to be effective," and thus "The initial practical question is about courses of action and not about themselves," or their motives1.988 And to this extent, we can reply to Watson in the same way as Pink:

---

985A full explanation of this issue, however, requires detailed treatment of the theme of authenticity in the writings of several major existentialists, and is beyond the scope of this dissertation, although I touched on this topic in the Introduction.


988Watson, p.109. This suggestion that identifications have strictly first-order motives seems to be made in part by way of response to Frankfurt’s notion that we can only identify with desires, we already have. I discussed this actualist restriction (AR) in Chapter IV §6.1.(A). Watson rightly rejects AR: "As I see it, agents frequently formulate values concerning alternatives they had not hitherto desired." It is easy to see why Watson (continued...)
even if ordinary practical deliberation begins by considering what ends different courses of first-order action may serve, identification lies in the background of such deliberation about particular acts, and is essentially concerned with evaluating the intrinsic qualities of motives, themselves, or the value of pursuing ends, rather than merely the ‘payoff’ value of achieving or realizing these ends through courses of action. Although preferences, desires, and intentions do not become authoritative or convey identification just because they are higher-iteration states about our first-order states, this does not mean that identification is not still an essentially reflexive attitude.989

Yet Watson’s view differs fundamentally from Pink’s, both because he still recognizes the phenomenological centrality of identification, and because he has a thicker conception of what it is to value actions, as opposed merely to desiring them. Watson distinguishes between desires grounded in an evaluation of the objective value of an end for well-being, and subjective preferences and inarticulate urges:

...to identify a desire or want simply in terms of its content is not to identify its source(s). It does not follow from my wanting to eat that I am hungry. I may want to eat because I want to be well-nourished; or because I am hungry; or because eating is a pleasant activity. This single desire may have three independent sources.990

These instances correspond to my own division between S3, S1, and S2 desires, respectively, except that I would regard the variations in Watson’s example as different desires with some real change in their content, rather than as an identical desire with three different sources. In

---

988(...continued)
seems to take this point —the falseness of AR— as reason to think that identification does not turn on agents asking themselves “which of their desires they want to be effective in action” (Ibid). Yet these are really separate issues: Watson can be right in holding AR false and wrong in grounding identification solely or mainly in first-order motives.

989In other words, the hierarchical feature of Frankfurt’s model thus seems to be phenomenologically valid, even if it does not provide a way of reducing intrapersonal identification (and/or alienation) to ordinal relations among motivational states like preferences, desires that do not have identification built into them.

general, however, for Watson values depend on "a person’s judgment of the good," whereas his appetites and passions do not.\textsuperscript{991} Thus an agent’s "valuational system" in Watson’s terms is similar in form to the evaluations which Stump believes determine our identifications: our system of values, combined with factual beliefs, "yields judgments of the form: the thing for me to do in these circumstances is, all things considered, a."\textsuperscript{992}

Watson admits that it is not easy to give a completely satisfactory account of such valuations. He points out that one may have inclinations that rest solely on one’s "acculturation rather than upon a current judgment of what one is to do:" yet he admits that such "acculturated attitudes may seem more akin to evaluation than to appetite," and may be articulated as judgments about the contrastive worth of options.\textsuperscript{993} In the end, he says our values are best defined according to our "more or less long-term aims and normative principles that we are willing to defend."\textsuperscript{994} Since this seems to be a much looser notion that the Thomist idea of eudaimonistic judgment, however, it is not easy to tell whether motivation by values would amount to S3-desires for Watson, or could include a wider set of motivational states grounded on considerations that do not ultimately turn on what contributes to the agent’s flourishing. Either way, however, Watson concludes that what distinguishes first-order desires with which we identify is not their endorsement by any higher-iteration state, but simply the fact that they "result from practical judgments;" unlike desires that result only from passions or appetites, they have the "special status" of agent-

\textsuperscript{991}Ibid, p.102.
\textsuperscript{992}Ibid, p.105.
\textsuperscript{993}Ibid. Watson also appears not to consider the problem that our strong evaluations themselves may simply be a result of conditioning or the pressures of very traditional expectations in our society. Thus someone may make judgments that we may think any unfettered deliberation guided by the natural light of human reason would have to reject, e.g. that having their daughter undergo ritual clitorectomy is the ‘the good to be done.’
\textsuperscript{994}Ibid.
authority.\textsuperscript{995}

In response to the concern that one might still "disassociate" oneself from evaluative judgments of actions and desires formed on their basis, however, Watson focuses on the idea of an overall "valuation system."\textsuperscript{996} Like Taylor, he holds that "One can dissociate oneself from one set of ends and principles only from the standpoint of another such set that one does not disclaim."\textsuperscript{997} And more importantly, like Taylor, he takes this hermeneutic inescapability of perspective in the will to imply that there must be some ends and principles that are ultimate for a person, or with which she is decisively identified. These form the center of an evaluational system that "may be said to constitute one’s standpoint, the point of view from which one judges the world."\textsuperscript{998} Thus —though Watson does not explicitly draw this conclusion— on his 1975 view, a desire, intention, or plan of action is \textit{id}entified\textit{f}ied with the person if it coheres with her overall valuational system. It is when a person is moved to act in a way (or on a motive\textsubscript{1}) which contradicts the ultimate standpoint from which she judges the worth of actions (and motives\textsubscript{1}) that she is ‘unfree’ through no external interference.\textsuperscript{999}

Like Taylor’s account, however, Watson’s early account does not seem to take seriously enough the \textit{irreducibly first-personal} aspect of identification. If I evaluate a certain course of action CA as ‘worthy of pursuing’ there remains the question: \textit{for whom}? I may think CA would be a very worthwhile course for \textit{you} to pursue in your circumstances (which are

\textsuperscript{995}Ibid, p.109.
\textsuperscript{996}Ibid, pp.105-6.
\textsuperscript{997}Ibid, p.106.
\textsuperscript{998}Ibid. I assume that by this, Watson means a \textit{practical}, not merely a speculative standpoint. In other words, this system defines the agent’s perspective, or determines what counts as authoritative for her.
\textsuperscript{999}Ibid, p.104: the distinction between Reason and Appetite is the basis for the possibility that "a person may be obstructed by her own will" (taking "will" here in Frankfurt’s initial sense as the motive, which can be a mere appetite, on which we in fact act).
similar to mine), and although this says something about me, my recommending it to you does not determine what is authoritative for my agency.

It is true that if one identifies with or alienates any first-order state, then one has an evaluational system with which one is automatically (or nonderivatively) identified: call this the Evaluative Principle. But without assurance of the antecedent, the consequent of this principle cannot be detached. Thus it may indeed be impossible to actively alienate one’s entire evaluational system —since having any positive identifications, depends on making some strong evaluations, as we have seen— but it may be possible to have an overall valuational system (even one that is quite coherently developed into a life plan\textsuperscript{1000}) while still remaining fundamentally neutral towards it or aloof from it. Identifying with a course of action thus always means something more than simply evaluating it as objectively worthwhile.

However, there is still a kernel of truth in Taylor’s and Watson’s arguments that we cannot identify with or alienate our evaluations ‘out of nothing,’ but must always do so from a perspective that already involves some evaluations and commitments. The truth this reflects is the fact that a person’s highest-order will, which conditions (though without determining) all her choice of reflexive attitude, can never be simply empty, or utterly wanton. Persons by nature always already ‘are’ some complete identification, or other. As I suggested at the end of the preceding chapter, a developed existential account of decisive identification would acknowledge that persons always have (or are) some highest-order volitional stance, so given the Evaluative Principle, this would appear to imply that they are never without some evaluational system.

The problem with this line of reasoning is that self-undermining highest-order volitions

\textsuperscript{1000}As Watson admits, the Rawlsian idea that "most people have articulate ‘conceptions of the good,’ coherent life-plans, systems of ends, and so on, is, of course, something of a fiction" (Ibid, p.105).
are possible, such as the will not to care about anything, or not to appropriate any valuational system at all. Thus it is not clear that decisive or complete identifications derive their authoritative quality simply from being our strongest evaluations: on the contrary, it seems likely that a person’s highest-order will can (and in the immature stages of life often may) consist simply of the tacit volition not to forms strong evaluations, to remain a ‘simple weigher’ by default, as it were. If he has inherited something like a valuational system from his culture, or even if he has developed one through his own reasoning, the agent may still decisively identify with holding back from engaging himself deeply in this system, or from taking it as the basis for his projections and willing that it motivate and guide his actions. Such a person shies away in general from developing inward or ‘deep’ character. Admittedly, he must do so from a ‘perspective’ that is absolutely and unqualifiedly his own, or that is the volitional center of ‘who he is,’ and that is not simply empty. But this does not mean that his ‘center’ consists in cognitive judgments that it is not good to take responsibility for himself or actively shape himself in accordance with ends it is worthwhile or meaningful to care about; on the contrary, it may be a purely volitional stance not controlled even by his highest evaluations. That is, it may be a pure rebellion against the call, nascent in his own personhood, to develop his capacity for an inner character; i.e. a resistance against the natural course or minimal teleology implied in the existential-hierarchical model of the person we have been developing.

Unlike Stump and Taylor, however, Watson later openly admits the fundamental difficulty for all the rational-evaluation analyses of identification:

Notoriously, judging good has no invariable connection with motivation, and one can fail to ‘identify’ with one’s evaluational judgments. One can in an important sense fail to value what one judges valuable. But even if this conflation is rectified by construing valuing as caring about something because (in as much as) it is deemed to be valuable....[w]hen it comes right down to it, I might fully ‘embrace’ a course of action I do not judge best; it may not be thought best, but it is fun, or
thrilling; one loves doing it, and it’s too bad it’s not also the best thing to do, but one goes for it without compunction. Perhaps in such a case one must see this thrilling thing as good, must value it, but again, one needn’t see it as expressing or even conforming to a general standpoint one would be prepared to defend.

It is possible to identify with courses of action (and presumably also motives) which do not fit with one’s valuational system; as Watson says, such cases may be perverse, but they are not for that reason necessarily "cases of compulsion" or even of "weakness of will." In this light, Watson acknowledges that, while its anti-Humean intention was sound, his 1975 account of identification was "altogether too rationalistic."

While S3 desires are qualitatively distinct from Humean motives, this is not sufficient to make them constitutive of volitional identification. What is needed, as Watson sees, is an account which is non-Humean in its recognition that identification is qualitatively distinct from an agent’s appetites and passions, but which also does not purport to explain this qualitative difference solely in terms of cognitive/judgmental bases, or ‘intellect-sources,’ and does not reduce identification to "brute self-assertion" either. Watson in the end professes not to have any satisfactory explanation for identification, and leaves the issue unresolved. However, my account in terms of projections is designed to meet precisely the conditions Watson recognizes. And as long as this account can ally predictable fears that it might ultimately rest identification in arbitrary fiats, this existentialist approach to

---

1001 The Thomist would of course insist on this, but unconvincingly I think: in the end, this claim is backed by nothing but the ideological force of Plato’s assumptions about motivation.


1003 Ibid. The last point of course depends on how we analyze "weakness of will." On some accounts, this may not require any "estrangement" from one’s actions. If, however, weakness of will requires that we act voluntarily and uncoercedly in ways that are contrary to our ‘inner character’ or deep cares, then a case where we identified unambiguously with an irrational or perverse project would not seem to be contrary to our cares, and thus not a case of volitional weakness per se. However, precisely how to diagnose what is going on in familiar examples may be quite controversial.

1004 Ibid, p.149-150.

explaining identification seems to have a better chance of success than either its emotivist or rationalist alternatives.

2.4. Velleman: Identification as rooted in the Desire to Act Rationally

In his remarkable book, *Practical Reflection* and subsequent paper, "What Happens When Someone Acts?" as well as recent presentations, J. David Velleman has developed a still more sophisticated rationalist account of identification that avoids some of the problems we have found for the accounts offered by Stump, Taylor, and Watson. Velleman’s account is rooted in his entire theory of agency, which is too rich and complex to be explained in its entirety here. Instead, in this section I will briefly evaluate the critique of Frankfurt that Velleman develops on the basis of his theory, and then criticize Velleman’s own proposed account of identification on this basis.

Velleman’s general approach is to argue against Davidson that reasons to act (or motivating factors such as beliefs and desires) do not directly cause the formation of intentions, nor do intentions directly cause corresponding bodily movements. Rather the "agent himself" must play two roles in action: "His role is to intervene between reasons and intention, and between intention and bodily movements, in each case guided by the one to produce the other." The standard Davidsonian belief-desire model of action fails to identify the mental states and psychological events which constitute the agent’s playing these two crucial roles, as Frankfurt’s addict-examples have evinced: "As Harry Frankfurt has


1008 Ibid, p.190. Note that for Frankfurt, *action* is defined only in terms of the second of these two conditions: we act when we guide our behavior in accordance with our intention. *Autonomous action* fulfills both conditions: we guide our intention in terms of a motive we select.
pointed out, an agent’s desires and beliefs can cause a corresponding intention despite him, and hence without his participation."\textsuperscript{1009} Velleman adds to this his own example in which an agent’s unconscious resentment of his erstwhile friend leads to a latent intention to break the friendship, and this in turn causes his voice to rise in reaction to comments during their conversation.\textsuperscript{1010} In this case, the agent is not \textit{active} in raising his voice or becoming angry (acting on the latent desire to break the friendship), \textit{not} because the motive’s causal effects on the intention or the intention’s causal effects on the behavior were somehow defective, but because the agent did not contribute to their effects in either stage. Such contribution is therefore not constituted by the motive and intention "manifesting their ordinary causal powers;" it is something more.\textsuperscript{1011}

Velleman rightly interprets "identification" as Frankfurt’s notion for this additional element: "what makes the difference between defective and full-blooded action" (in which the agent "participates") is that "in the case of the latter, the agent identifies with the motives that actuate him."\textsuperscript{1012} But Velleman wrongly supposes that for Frankfurt, the agent’s identifying with a motive\textsubscript{1} means "adding to the force" of the motive\textsubscript{1} in a \textit{causal} sense, or "throwing his weight behind [it]" in a way that adds to the causal factors producing an intention to act.\textsuperscript{1013} As we have seen, identification is not a causal concept for Frankfurt, and this will prove important.

Velleman argues that Frankfurt and his interlocutors have not been able to provide a

\textsuperscript{1009}Ibid, p.191. He notes that in "deviant causal chains," an intention may also cause a behavior without that intention guiding that behavior. Frankfurt also makes this point in "The Problem of Action" (although Velleman cites Bishop, Harman, Peacocke, Taylor, and Goldman on this point, he leaves out Frankfurt).
\textsuperscript{1010}Ibid, p.192 (also see his explanation of this case on p.198).
\textsuperscript{1011}Ibid, p.192.
\textsuperscript{1012}Ibid, p.199.
\textsuperscript{1013}Ibid, p.199. Also see p.201.
convincing reductive analysis of identification, or (in Velleman’s terms) an account of the
states which constitute an agent’s participation in forming intentions based on motives.
Second-order desires fail for the reasons we have seen,\textsuperscript{1014} and Watson’s evaluational system
fares no better because "a person can be alienated from his values too; and he can be
alienated from them even as they continue to grip him and to influence his behavior."\textsuperscript{1015} As
we have seen, this question is more complex than Velleman here acknowledges, but his
verdict on it is substantially correct. Velleman also critiques Frankfurt’s account in terms of
decisive commitments: he claims that his example (above) shows that decisions can arise in
us unwittingly (as opposed to our \textit{making} them), and thus "be foreign to the person in whom
they arise."\textsuperscript{1016} Our previous analysis requires resistance to Velleman on this particular
point, since his claim rests on the dubious notion that there can be unconscious decisions. It
seems possible to hold with Frankfurt that conscious decisions, at least, are inalienable. And
this does not seem to be, as Velleman suggests, just because decision is \textit{defined} in a way that
makes implicit reference to our participation or identification.\textsuperscript{1017} Conscious decisions form a
set of experienced phenomena that seems to be both distinct from and more basic than our
experiences of identification.

This point aside, however, Velleman is right that decision cannot be the whole
explanation of identification, and that we cannot simply leave this phenomenon as primitive
or unanalyzed.\textsuperscript{1018} He also rightly sees that for Frankfurt, nothing can qualify as an

\textsuperscript{1014}Ibid, p.199.
\textsuperscript{1015}Ibid, p.200.
\textsuperscript{1016}Ibid, p.200.
\textsuperscript{1017}Ibid, pp.200-1, note 27.
\textsuperscript{1018}Ibid, p.201.
identification unless the agent cannot "dissociate himself" from it.  

But for this reason, Velleman says that assuming that identification (and the self-identification it automatically involves) occurs is just to assume that agent-causation occurs, "which is what we set out to show in the first place" against the Davidsonians. To avoid begging this question, we must reduce identification to other less potent mental phenomena "whose existence we can assume without presupposing that agent-causation occurs."  

There are two closely connected mistakes behind this conclusion. First, Velleman wrongly assumes that identifying with a state refers to the agent's (or what constitutes him) playing a causal role in bringing about that state: for he writes that an identification theory answers the question about "how an agent causally contributes to the production of his behavior" only if on that theory, "identifying with motives entails somehow making a causal contribution to their operation."  

Second, Velleman assumes that the theoretical point of analyzing identification into more primitive terms is to show that it is plausible that agent-causation in this sense occurs.  

The second error is the more fundamental one: as the argument of Chapter IV suggests, the existence of identification —or at least its indispensibility for our conception of ourselves as agents— is not in serious doubt.  

The task of a theoretical explanation of identification

---

1019 Ibid, p.201.
1022 Ibid, p.201.
1023 At one point, Velleman himself admits as much: "What makes us agents rather than mere subjects of behavior—in our conception of ourselves, at least, if not in reality—is our perceived capacity to interpose ourselves into the course of events in such a way that the behavioural outcome is traceable directly to us" (p.193). If this is what he means by agent-causation, however, it is unclear why Velleman should think that its (continued...)
should be to explain as convincingly as possible phenomena that are already apparent to us, where what counts as ‘convincing’ is guided only by the shape of the *phenomena themselves* (rather than by whether the result thus purely derived fits someone’s preconceived metaphysical biases about what sorts of entities should be admitted into the world). Failure to observe this point makes possible the first error of assuming that identification must be a form of agent-causation. For if we start out simply trying to explain the phenomena as we find them, we may not (as Frankfurt *was not*) be led to an account of identification that links it to the causation of action.

It is important to recognize these errors, because in effect they lead Velleman to the conclusion that only an analysis of identification which explains it in terms of states that constitute agent-causation could meet the Qualitative Condition, which misleads him into thinking that the concept of identification is otiose. For Velleman the real problem becomes to figure out how agent-causation is constituted; relative to this, the phenomena of identification are secondary. Thus finding states that are “functionally identical to the agent,” i.e. states in virtue of which a person plays the roles distinctive of agents (intervening causally in motive-intention connections), becomes replaces the analysis of identification.

---

1023(...continued)

1024I have put this as strongly as possible in order to offset the naturalizing tendency in Velleman’s approach. For more discussion, see my brief defense of the phenomenological method at the end of the Preface.

1025Ibid, p.202: if we explain identification by picking out "mental states and events in terms that do not presuppose any causal contribution from the agent, then we shall have picked out states and events from which the agent can in principle dissociate himself" (i.e. that fail to meet the Qualitative Condition).

1026Ibid: “Since self-identification won’t serve our purpose unless it’s conceived as something to which the agent contributes, rather than something that happens to him” [Velleman’s equivalent for satisfying the Qualitative Condition], and since this means agent-caused action, "reducing self-identification to mere events and states is unlikely to be any easier than reducing action itself” (p.203).

1027However, note that Velleman is right when he says that "identifying with something" in Frankfurt’s sense "is a relation that one bears to something functionally distinct from oneself," whereas whatever identification itself amounts to is functionally identical to the agent (p.204). This is just another way of putting the point made in Ch. IV that identification is *conveyed* to states with which we are not automatically identified, while we (continued...
So conceived, Velleman argues, the agent’s role is played by whatever attitude(s) or motive(s) are always at work in critical reflection on motives, but which never themselves undergo scrutiny (and therefore can never be rejected in the critical evaluation). This, then, "can only be a motive that drives practical thought itself," or that always motivates reflection on motives at any level.\textsuperscript{1028} And this fundamental drive is "a concern for acting in accordance with reasons." What is thrown behind motives with which we identify is "the additional motivating force of the desire to act in accordance with reasons."\textsuperscript{1029} This desire is understood as a broad type that could be instantiated by several sub-types. What it requires in general is that our actions be intelligible to us as having been motivated in a way that "makes sense" to us, or that is rational, or that had the strongest reasons on its side. As Denise Meyerson explains, on this account:

I have contributed to my behavior when the desire to act in accordance with reasons has intervened between my motives and behavior, adding its motivational force to a motive which appears to provide the strongest reason for acting.\textsuperscript{1030}

Against this approach, Meyerson raises the objection that a compulsive person’s action might be overdetermined, caused both by her positive rational assessment of the motive on which she acts, and by the compulsive force of that motive. And she rejects the idea that such a person could simply be identified with her act though her addiction makes her incapable of acting otherwise.\textsuperscript{1031} In other words, she seems to hold that Velleman’s account

\begin{footnotes}
\item[1027] (…continued)
\item[1028] Ibid, p.206.
\item[1029] Ibid, p.207.
\item[1030] Denise Meyerson, "When are my Actions Due to Me?," \textit{Analysis} \textbf{54.3} (July, 1994): 171-174, p.171.
\item[1031] Ibid, pp.173-4. Yet Meyerson mistakenly thinks that for Frankfurt, a willing addict is identified with his act because his desire to act on the addictive desire is a partial cause of his action (p.174). As we have seen, the causal role of the higher-order volition is not what matters for Frankfurt, and in addition, such a volition (continued…)
\end{footnotes}
is plausible only for cases in which the extra motivation provided by the higher-order assessment of the motive, on which we act is what makes the difference in our so acting.

But it seems that Velleman could modify his analysis of identification to satisfy Meyerson’s concern without giving up its central claim, and without giving up the Frankfurtian idea that we can be identified with actions we could not avoid doing (or with motives on which we could not avoid acting). Suppose we defined identification with an action in terms of its flowing from a type of reasons-responsive mechanism in Fischer’s sense, with the modification that what this mechanism evaluates is primarily the value and strength of reasons behind different possible motives, on which we might act. An agent (like the willing addict) might still act on the same motive, if it were not rational in her assessment to act on this motive, and in this counterfactual sequence, her action would thus not flow from a process that is higher-order reasons-responsive (as we might put it). But in the actual sequence, her desire to act on reasons produces a reasons-responsive evaluation of her different motives, and this evaluation plays a causal role in her acting on a given motive: she is identified with her action. This shows how Velleman’s approach could be recast in a Fischerian actual-sequence model of identification that focuses on the reasons-responsiveness and causal efficacy of higher-order evaluations of motives; we are identified with an action when our positive evaluation of the reasons for acting on a given motive plays a causal role in our acting on . To be identified with the motive on which we act in this model does not require that in counterfactual sequences where we do not identify with it, we do not act the same way: thus identification need not exercise what Fischer calls "regulative control" over our actions.

The real problem with Velleman’s positive account is more subtle. It avoids the...
difficulty faced by other rationalistic accounts because it is framed in terms of a general 
*desire* for rational action. Velleman says that a person might "disown" this desire, but only 
by giving up acting as an agent (or in Frankfurt’s terms, becoming wanton).\(^{1032}\) But is this 
supposed concern for the rational justification of our motives properly described as a 
"desire" in any *substantive* sense of desiderative motivation? It is if we limit ourselves to the 
kinds of reasons or considerations that underlie S2 or S3 desires: e.g. that we simply find 
something preferrable, or that we think something is objectively good for us or ‘completes’ 
our well-being in some fashion. We may be said to have a *generic* desire to act in pursuit of 
whatever we prefer, and a *generic* desire for our good conceived in the fullest objective 
sense. A given motivation towards an end may be strengthened by the force of these desires 
if it fits their criteria. For example, if I am drawn to orange juice because I like its tangy 
taste, my particular S2-motivation to drink it is reinforced by judgment that orange juice is 
also good for me (gives me vitamins I lack), which produces a concurrent S3-desire to drink 
the glass of juice in front of me. We could also have judgments that give us an S3-desire to 
cultivate such an S2-desire (supposing that our liking for the taste is somewhat weak), and in 
that sense be adding to the motivation to drink the juice. But none of this constitutes 
*identification*, since we can alienate our S3-desires as well as our S2-desires without ceasing 
to operate as an agent. So these generic desires do not play the role Velleman is after.

Instead, as I argued in Chapter IV, we discover inalienable motivations\(_2\) to act on 
motives\(_1\) only when we consider agents who *project* for themselves the goal of developing 
and acting on particular motives\(_1\) over time. Different types of reasons relating to qualities 
of the motives\(_1\) at issue may ground such projections, but it would be deceptive to speak of 
the projective motivation\(_2\) in these cases as a "desire" to act on such reasons: for these are 

\(^{1032}\)Velleman, "What Happens When Someone Acts?", p.208.
Since like Frankfurt, we might want to define "action" and "agency" according to a less demanding baseline that allows non-human animals and young children to perform acts and be agents in this sense. This of course is not a "nature" or telos in the eudaimonistic sense. It may constitute the most developed or most aesthetically and ethically appropriate form of motivation for persons, without this meaning that a person who conforms to it is necessarily happiest or has the greatest well-being.


precisely the sorts of reasons that do not ground *desires* in any substantive sense of the term. It would be more accurate instead to speak of a generic *project* of projecting motives for oneself on grounds that one can understand: perhaps this is indeed the *minimum* fundamental project without which a person fails to qualify as an agent (at least in the personal sense). A much-enriched version of this fundamental project might define a higher or more mature level of agency, or a *normative ideal* for personhood: the project of becoming motivated in ways that are challenging, interesting, varied, morally acceptable, ethically inspiring (aimed at bringing good into the world in general) and sufficiently unified to give life a meaningful shape. As we saw in the Introduction, this is certainly *more* than is required to count as a personal agent, but it is the ideal towards which the projective character of our volitional capacities seems to point.

Only when Velleman's "desire to act in accordance with reasons" or on justified motives is translated into this projective analogue does it seem to capture a kind of motivation that is inalienable (or inalienable without abandoning personal agency) and that lends its force to other motives. Note that if it is understood in purely motivational terms (which need not also be causal terms), then Velleman’s notion that when we identify with a motive we are reinforcing it "by another motive" fits with my analysis. The motivation by which we strengthen or reinforce motives, when we identify with them is, however, a *projective* rather than desiderative one. It is not the force of *reason* but rather the kind of reasons —reasons to project motives,— that makes this motivation identical with the agent’s role. Moreover,

---

1033 Since like Frankfurt, we might want to define "action" and "agency" according to a less demanding baseline that allows non-human animals and young children to perform acts and be agents in this sense.

1034 This of course is not a "nature" or telos in the eudaimonistic sense. It may constitute the most developed or most aesthetically and ethically appropriate form of motivation for persons, without this meaning that a person who conforms to it is necessarily happiest or has the greatest well-being.

the weight we "throw" behind our incorporated motive, is not necessarily causal. Hence the motive, and our projection of it together need not "form the strongest combination of motives" in the sense of successfully determining our action, as Velleman implies; it is only the authoritative combination. So Velleman was very much on the right track: he went wrong only in defining the analysis of identification too narrowly as a causal investigation, and in lacking the concept of projective motivation in working out his own alternative to Frankfurt’s analysis.

V.3. Decisive Identification as Self-Discovery: Bransen and Tugendhat

In view of the regress and ab initio problems for Frankfurt’s hierarchical account of identification, Jan Bransen follows Watson in proposing that we leave out hierarchy entirely in explaining identification. But unlike Watson, Taylor, Stump, Velleman and others, Bransen does not construe the agent-authority of states as constituted by their flowing from some kind of rational or evaluative source. Instead he follows Ernst Tugendhat’s neo-Heideggerian notion that autonomy or self-determination in the strong sense depends on a process of self-interpretation in which we try to answer the fundamental practical questions of what kind of life we should lead, or what kind of person to be.

As Taylor explains in a review, for Tugendhat, as for the existentialists, "There is in our universe of discourse something we want to refer to as a self, only because we are the kinds

---

1036Ibid, p.209. The problem with this of course is that it rules out identification that is ineffective in controlling action. This is another problem for Velleman’s account, for the unwilling addict surely identifies with a motive that, even including her identification with it, is not the strongest combination of motives in the causal sense.

1037A fuller account of rationalist options would include a discussion of Susan Wolf’s account of autonomy.

1038Ernst Tugendhat, Self-Consciousness and Self-Determination, tr. Paul Stern (MIT Press, 1986), esp. chapters 8-10 on Heidegger and the "Relation to Oneself." Also see the translator’s summary, pp.xii-xvii. Note that what kind of person we should be (morally speaking) is part, but surely not the whole, of this question.
of beings whose being is in question, for whom the issue arises: what kind of being to be?" But Tugendhat believes that when we raise this practical question in its "comprehensive sense" as related to our "fundamental life-possibilities"—which corresponds to what Heidegger calls the "for-the-sake-of-which" of Dasein, or what Frankfurt calls decisive identification, or what Williams calls our ground projects, or what Taylor calls "radical self-reflection"—such ultimate commitments must be determined through practical deliberation "directed to an objectively justifiable choice of what is good," or in light of criteria and standards for a good life and character that allow for criticism of our inward stand by others. This is true even if our ultimate commitments or comprehensive stand cannot be determined solely on "objective, rational grounds," as the rationalist theories we have examined imply. Thus as Stern puts it,

"...Tugendhat’s conception of the reflective self-relation represents a middle position between the Kantian idea of the autonomy of reason and the Heideggerian account of authentic choice. In his view, autonomous choice requires that one decide in the mode of rational volition. This means that the agent’s choice of how he wants to live must be governed by rational considerations insofar as these are relevant to the determination of his true good, but it must also reflect a distinctively individual exercise of preference for which the agent must bear sole responsibility....the fundamental Kantian intuition that autonomy can be attained only by means of a relation to reason and truth is sustained, but the standards of reason and truth are no longer severed from the irreducibly individual context of choice within which we decide how we want to live."

As this already makes evident—and as a more detailed discussion of Tugendhat’s own rich analysis would confirm—his ‘middle view’ of decisive identification converges

---

1042 Ibid, p.xv.
1043 Ibid, p.xvi.
1044 Ibid, p.xvii. Stern goes on to give a very helpful analysis of the similarities and differences between Tugendhat’s views and those of Frankfurt, Watson, and Taylor (pp.xvii-xxx).
substantially with my own account in terms of second-order projective motivation. What the projection-account adds is (a) a clearer sense of how identification can be both a volitional expression of the individual and yet grounded in reasons; and (b) a sense of how the considerations that ground identifications can both underdetermine them and not turn ultimately on the agent’s concern for her own individual good. In the latter respect, the projective account inverts the implication of eudaimonistic explanations that trace volitional identification to S3-desires.

Bransen extends Tugendhat’s approach by arguing that Frankfurt’s later account of autonomy in terms of decisive identification (in "Identification and Wholeheartedness") is still in danger of sounding too decisionistic. It could be interpreted as what I will call the Simple Self-Assertion (SSA) model of identification, according to which an act I of identification with a desire D conveys the authority of the agent to D because "that is precisely the intention" of act I.1045 As Bransen says, such an SSA account lacks explanatory force, because "on this interpretation the act of identification is nothing but an arbitrary act of transferring the authority of the agent to the desire in question."1046 This would be equivalent to what Watson and Velleman call a conception of identification as a sui generis state or psychological primitive. On another interpretation, which I’ll call the Comparative Self-Illumination model (CSI), when you identify yourself with a particular state, S, you are comparing "yourself" and S as objects, and recognizing that S sheds light on yourself. On this reading, to identify yourself with a desire to avoid cigarettes, for example, is to describe the fact that this desire is an important part of "yourself," or that this desire has a "dominant,

influential, or otherwise significant presence in your inclinations and deliberations." But as Bransen says, this is descriptive sense of "identify," and thus the CSI model does not explain how we endow particular states with authoritative force or importance, or direct the inward "self" with which we compare them.

Instead Bransen suggests that identification is constituted by a special kind of self-interpretation, in which we identify ourself as one sort of person or another: "it becomes clear to a person that she is really or most truly, motivated by this or that desire." There is a kind of middle-voiced ambiguity in the sort of self-interpretation Bransen has in mind: it is not simply a matter of describing the well-defined identity and commitments one has, nor of determining oneself by mere fiat, but rather of becoming responsive to a latent volitional orientation and thus of discovering "the identity of an entity that at first appears to be undetermined." He adds:

...at first you are confronted (as we say) by yourself: a person whose identity, as far as its true motivations are concerned, is concealed, or undetermined. But then something happens, something that can be understood in both a passive and an active sense: you succeed in determining or articulating what turns out to be your proper identity by realizing that you are [in this case] the person moved by the desire not to touch any cigarettes. This is not merely an arbitrary decision on your side....the person in doubt [must be] given in such a way as to enable the agent to discover whether or not identifying the person in doubt as the motivated person implies an increased understanding, or a more accurate articulation, of the identity


1048 Still, this does not justify Bransen’s conclusion that an act of identification is not "an act of identification with" some other state at all (p.7). Bransen is right, as we have seen, that the fact that a state conveys identification to other states is not what makes it an identification: rather, its identification-conferring power is derived from its being identified with the agent in another non-conferred sense, which is what Bransen means by "identification [of the agent] as..." But states that are directly identified with the agent as an expression of (or aspect of) her inner self, are still states that are about the agent’s motives and dispositions, though again this reflexivity is not what constitutes their authoritative quality. In that case, they will still be states of identification-with. Bransen sees this later: "any thought about any alternative of oneself will be a higher-order mental state, but this fact has nothing to do with whether or not a particular alternative is considered to be the right or true [i.e. authoritative] one" for us (p.10).

1049 Bransen, p.8.

of this person in doubt. However, it is also not merely a matter of ‘wait and see.’ The act of identification as is an act, an attempt to articulate your proper identity, and attempt to interpret yourself.1051

There is something appealing about this account. Note, as we saw earlier, that Frankfurt’s revised analysis of decisive identification or wholeheartedness does not mean either the SSA or CSI models. Rather he understands decisive identification in terms of volitional necessities (IV.5.2). As Frankfurt imagines this, when we experience such necessities, which are often only discovered at moments of crisis in our lives, our true identities become apparent or clear to us in a way not previously accessible to us. I think this does fit with one aspect of our experience of identification; namely, that we may not subjectively prefer or objectively approve of the devotions and commitments we discover at the root of our ‘selves,’ although we must recognize them as ours. Bransen’s notion of self-interpretation preserves this insight in Frankfurt’s approach in a model that is less passive than the volitional necessity account of identification, because Bransen suggests that we actively undertake to articulate our more hidden commitments, rather than waiting for events to clarify them, or for crises to force us to take a stand.1052 And in this respect, as he says, Bransen’s approach is in harmony with the hermeneutic tradition.1053

But in the same way as Frankfurt’s notion of volitional necessities, this account implies that there is just a fixed truth about our inner character that guides us when we actively seek to articulate our identifications. And this seems less phenomenologically adequate or convincing related to our actual experience. It is true that the agent’s relation to her most

1051Ibid, p.9.
1052For example, Heidegger thinks that in the face of an authentic or unvarnished appreciation of our own mortality and finitude, we may be empowered to see more clearly what we have been living for, and thus to ask (often for the first time in life) whether this is really what we want or should be living for. But this also implies an ability to change our ultimate commitments if we find them inferior or ignoble.
inward level of volition, or highest-order cares, must in part be a matter of self-discovery and improved self-understanding about a character that is not instantly changeable, and that in some respects is perhaps close to fixed. But such self-interpretation itself is important to us not simply because we want to know the truth about ourselves —however dark or glorious it is— but also because we care about the evaluation of our highest-order will. And this implies the ability to change. Our ultimate identifications are not simply discovered, but also projected; it is perforce not easy to alter old commitments or establish new ends concerning ourselves at this most fundamental level, but it is never impossible for us: the highest-order self is never a fixed monad, but always a sphere of more-and-less possible highest-order commitments around the actual identifications, we are factically in any duration of time (note how this fits with the description at IV.6.1-c). It is this modal extendness which allows comprehensive or ultimate self-interpretation still to be critical: we not only can consider the de facto highest-order reflexive ends we have; we can also compare their de jure values to the values of other volitionally possible highest-order commitments within our horizon, and choose or project new ones out of this range for ourselves, in light of this evaluation. This more active determination rather than mere discovery of our identifications, however, is still not an arbitrary fiat. It is guided, but not simply by our factical identifications; rather, it is also guided by the evaluation of our possibilities, and the grounds we can discover for projecting some of them rather than others. These grounds are never themselves completely grasped in reflection, and may not determine a unique choice, or may leave the relative importance of the relevant considerations themselves to some extent undetermined until we have projected highest-order ends for ourselves, but this is precisely why the initiating power of projection is required for us to be authentically in control of our highest-order self. And this is surely why Tugendhat
recognizes an "irreducibly individualizing, volitional component" in his conception of our practical self-relation.

V.4. **Buss’s Nonattitudinal Account of Identification**

In a remarkable paper, Sarah Buss defends the idea that agents can act intentionally and yet nonautonomously, but argues that this distinction "has nothing essential to do with an agent’s *attitudes* towards her action, and in particular, nothing essential to do with whether the agent endorses, identifies with, or is in any way ‘satisfied' with her action."\(^{1054}\) In my terms, Buss holds that the autonomy of an intention does not depend on any *weakly higher-order* attitudes about it: rather, for Buss, the intention has authority as an expression of the agent herself *unless* her all-things-considered preference for this particular act-token results from nonrational influences that are neither "a function of her character" nor "compatible with her living a good life."\(^{1055}\) While on Buss’s account, "No one...can be prevented from acting autonomously by being forced to act as she prefers *not* to act, all things considered," it is possible for "her preferences themselves [to] be the products of autonomy-undermining influences."\(^{1056}\) For example, someone who is "the victim of brutal rape" may act intentionally on her preference to cooperate without acting autonomously, since "she is so overwhelmed with fear that she cannot give serious consideration to the possibility of fighting back."\(^{1057}\) As Buss clarifies, this does not amount to "claiming that a person is to be identified with her reason alone" as opposed to her emotions, since nonreasons may be part

---


\(^{1055}\)Ibid, p.96 (note that the page number on this page is misprinted as ‘94’ in the journal).

\(^{1056}\)Ibid, p.105.

\(^{1057}\)Ibid, p.106.
of the agent’s "true self," and may even facilitate practical reasoning. Rather, to defeat autonomy, nonrational influences must "sabotage" her practical reasoning by being both unhealthy (not "compatible with human flourishing if it regularly made itself felt under similar circumstances") and not "constitutive of the agent’s character." For example, though two agents may both suffer from anger intense enough to be unhealthy, "Whereas the irascible agent’s anger is a form of self-expression, the mild-tempered agent’s anger is something that ‘happens to’ her.”

As Buss notes, this account is similar to Wolf’s in that on both, "a person’s accountability for her behavior is a function of her relationship to ‘the good,’ and thus autonomy depends on what Benson would call substantive rather than purely procedural conditions on the process of intention-formation." But in contrast to Wolf’s account, Buss’s requirement that nonautonomous influences be ‘out of character’ for the agent means that "a person can act autonomously even though she has an evil character, and is incapable of doing the right thing for the right reasons..." This means that her account does not fall prey to the same criticism Benson levels against Wolf’s: "it is paradoxical to hold that free agency necessarily eludes those whose normative capacities are, by Wolf’s lights, insufficiently objectivist."

---

1061 Benson, “Free Agency and Self-Worth.” Both standard "structural and historical” accounts are procedural in the sense that "all these theories hold that the privileged volitional territory can be marked out independently of the content of the agent’s attitudes and that the unimpeded regulation of the agent’s effective motives and behavior by the appropriate region will guarantee free agency" (p.655). Buss’s approach would count as historical, like Fischer’s, since it focuses "on the process through which the agent acquired her effective motives" (p.654), but not as content-neutral, since it depends on the role of nonrational influences that are contrary to the human good in that process. Yet Buss’s conditions are much stronger than Benson’s requirement that the process be such as not to undermine the agent’s sense of her status as worthy to exercise her capacities.

1063 Benson, “Free Agency and Self-Worth,” p.655. Benson goes on to admit that a person who is "unable to
Yet I think Buss’s positive account succumbs to difficulties of its own. For one thing, it seems to treat similar cases unequally. Consider a miser who, because he is suddenly influenced by non-rational pleasure at the sight of an innocent young baby, gives her homeless mother a thousand dollars. Although he acts out of character, the non-rational inclination which influences him is (presumably) one whose regular influence on an agent’s reasoning would be compatible with the human good, and so his act is autonomous. But why, then, does the generous person who, when faced with the same homeless mother’s plea for help, acts miserly because of a severe stomach pain (whose regular influence is not compatible with flourishing) not act autonomously? She acts out of character, but so does the miser. It would seem that both are similarly situated relative to the nonrational influences on the formation of their intentions or all-things-considered preferences, so why should they be evaluated differently?

Buss may be able to elaborate her account to handle this problem. But the more serious difficulties concern her understanding of ‘character.’ Despite her caveat that acting ‘in character’ may not always be sufficient for an agent’s act to be autonomous —e.g. it is insufficient if she was "systematically discouraged from acquiring alternative character-traits"— it is questionable how on this account we could form our motivational character autonomously. For to effect changes in our first-order dispositions to (say) act on greedy motives, we must at first act ‘out of’ the character we are trying to change. The unwilling
addict is someone who will try, if possible, to change her dispositions and first-order motives in just this way. But these efforts and intentions are not to be explained by the chance interference of desires, or inclinations, that do not fit the pattern of her usually dominant motives, as when the miser suddenly has a temporary fit of generosity. These character-changing efforts have motives that are ‘out of character’ in quite another sense. To understand this requires, I believe, recognizing that there are two relevant ‘orders’ of character itself (in the thick sense of order as level of authority): our inward character consists in our authoritative will to develop and/or maintain patterns of (or dispositions to) certain first-order motives (rather than others) as the ones we regularly exhibit in action. The persisting will to alter or to maintain one’s outward character or dominant dispositions of first-order motivation must be understood as a state of inward character in this sense. Although the topic arose in recent analytic philosophy in the context of discussions about what it is for particular acts to be autonomous, identification is really a phenomenon of inward character in this sense, and its analysis should focus on how we become responsible for our character (in both senses), not simply on when particular act-tokens are imputable, or ‘free’ in some other sense distinct from simply being intentional. Thus for example, a willing miser who identifies with his disposition to act on greedy motives is still responsible for that character when he acts on a random generous motive, instead, while a genuinely unwilling miser has diminished responsibility for her greedy character, even when she acts in accordance with it, and even though these dispositions may not have been formed by systematic conditioning or pressures that would excuse her on Buss’s account.

1065 It seems to me that Buss’s analysis really starts from disagreement with Frankfurt’s account of coercion, and addresses the issue of identification from that point of view. Contra Frankfurt, she holds that people cannot be coerced in acting on motives, which they don’t endorse: if they act intentionally at all, they are coerced only if the preference underlying their intention was formed in the wrong way (i.e. by irrational influences that were out of character). But the issue of identification is inherently broader than the question of when first-order acts are involuntary or coerced, and likewise, though they are related, autonomy is not identical with voluntariness.
Rather than further evaluate Buss’s own account, however, let me turn to her critique of Frankfurt’s account. Buss’s critique has two related parts, both of which I think can be accommodated within an account of autonomy in terms of identification. The first is simple: Buss suggests that Frankfurt’s account is inadequate because, an ideology which someone wholeheartedly endorses can undermine her autonomy because it can influence her choices in a nonrational manner that is incompatible with human flourishing.

Although this is phrased as an objection to Frankfurt’s account of decisive identification, it can be interpreted as a version of the Origin of Identification objection (B1.ii): it concerns the question of whether the process which generates the states constituting any kind of identification must meet certain substantive conditions for them to constitute identifications, or expressions of the agent’s real self. As I’ve already argued, it is not clear that the conditions Buss singles out are the right ones, since they seem to involve the Aristotelian assumption that all agents necessarily identify with the motive to pursue their flourishing or well-being. However, as we have also seen, the possibility that some such substantive conditions on the originating process are essential to identification (or part of what qualitatively differentiates identifying states from mere second-iteration desires) is compatible with the basic idea that identifying states are still about first-order motives or dispositions (or ‘higher-order’ in the weak sense), even if this is not what makes them authoritative.

The second part of Buss’s critique is more complex. She argues that there is "an important and obvious sense in which no one can at t prefer not do what at t she intends to

---

1066 An adequate discussion of her model would require a treatment of her allied and similarly novel account in her recent article, “Weakness of Will,” Pacific Philosophical Quarterly 78 (1997): 13-44. But this is beyond the purview of my present purposes.

do." If one’s all-things-considered preference is to remain in the house, then one’s intention will be to stay inside, unless this preference is altered by other considerations.1068

"Accordingly, no agent sets herself the goal of doing what she prefers not to do, all things considered."1069 Note that this assumes something like the Davidsonian practical judgment model of intention-formation: it is deliberative rather than executive.1070 Whether this is right or not, however, it also seems that in this sense, an unwilling agoraphobic must have an all-things-considered preference to remain indoors, although she may not identify with the intention to remain indoors that issues from the preference. If so, then this model of intention-formation poses no challenge to Frankfurt.

But Buss thinks this appearance is deceiving. She agrees that we can prefer (all things considered) to perform a particular act (or act-token) in the situation, while not preferring to perform this type of act, because we’d rather not be in the situation where it is our best option at all: e.g. a person may engage in painful exercise only to stave off a degenerative disease, and thus prefer these particular acts of exercising, though she does not "endorse" this type of act, since it conflicts with ideals she pursued before the disease, or hopes she had concerning how to spend her time or devote her energies. Thus, "without failing to act autonomously, we can do things which, in an obvious and important sense, we would much rather refrain from doing."1071 This leads to a somewhat Sartrean critique of Frankfurt’s

1068Ibid, p.97. Like Kant, Buss says that "Whereas a person finds herself with certain inclinations, her preferences are the product of her own comparative evaluations." They involve judgment about what to do, all things considered (pp.97-98).

1069Ibid, p.97.

1070In particular, Buss says that if an agent takes herself to have insufficient reason to do some act, but behaves this way anyway, she is not acting at all: "her behavior is merely the unintended effect of some motivating force at odds with her reasons" (p.98). But then Buss’s model of intentions faces the same objections that Pink raises against Davidson’s "pro attitude model" (though in the respects that Pink shares Davidson’s theory of the motives for decisions, Buss would seem to agree with him—see IV §3.1 above).

1071Ibid, p.100. And she adds that this is possible without weakness of will, as well.
unwilling addict case.\textsuperscript{1072} The addict cannot \textit{really} be unwilling, she says, because when "this addict intentionally injects drugs into his veins, he prefers to do so in the only sense in which preferences are relevant to his autonomy."\textsuperscript{1073} He might not prefer to perform this \textit{type} of act, because unlike the willing addict, he is not "contented" with the desire which limits his options such that he must prefer the act-token of taking the drugs, all things considered.\textsuperscript{1074} But this does not mean that he was \textit{coerced} or acted nonautonomously when he last injected himself with heroin: rather, "he regards his addiction as an unfortunate, irremediable circumstance"\textsuperscript{1075} and in fact he takes the drugs to avoid losing autonomy by descending into a state where dire craving compels him to completely involuntary or unintentional behavior.\textsuperscript{1076} Thus his position is similar to the person with the degenerative disease, who cannot (without self-deception) believe that she is helplessly compelled because she doesn’t endorse the desire to exercise.\textsuperscript{1077} the only difference is that his options are limited by his desires, while hers are limited by external natural causes (the disease).

One immediate response to this argument might be that identification is not with acts, or their intentions, (or the preference so to act) but rather with the motive for the intention essential to the act (since the same act-intention may be formed on different motives). And

\textsuperscript{1072} Buss compares and contrasts her account and Sartre’s, pp.102-104, and notes 31-32 (p.119). While she resists his idea that someone who believes he is coerced by his desires is in "bad faith" (p.103), Buss agrees with Sartre’s insistence "that no person who does something intentionally can credibly protest that she really intended to act otherwise" (p.102), since someone cannot perform an act-token intentionally without "concluding that she has sufficient reason for acting this way" (p.118, note 31). An adequate response requires a full analysis of Sartre’s account of bad faith, but such an analysis will show that: (1) Sartre’s account is \textit{unrealistic} and denies people an ‘excuse’ \textit{too} vehemently in just this respect, precisely because he lacks the Frankfurtian notion that we may authentically identify with motives contrary to those on which we act; and (2) nevertheless, his own account of the phenomena of bad faith is in need of this notion of identification (see my unpublished paper, "Identification and Bad Faith: Combining Frankfurtian and Sartrean Insights").

\textsuperscript{1073} Ibid, p.101.
\textsuperscript{1074} Ibid, p.102.
\textsuperscript{1075} Ibid, p.102.
\textsuperscript{1076} Ibid, p.103.
\textsuperscript{1077} Ibid, p.100.
since the motive has to do with the type of act it is, whether we identify with this type is relevant for autonomy: although "we cannot be passive with respect to the what of our acts" if we executed this act-token, we can still be passive with respect to the why. As Velleman argued, the agent intervenes between his motive and the intention formed on its basis. But this is not enough to address the central point of Buss’s critique, which draws on her parallel between the unwilling addict and other persons whose best act-options are ones they have reason to be unhappy about, but which they still evidently perform ‘autonomously’ (or without coercion).

To see why this objection fails, we have to recall Frankfurt’s conclusion in "Identification and Externality" that identification cannot be constituted by simple weighing of alternate possible motives, or by preferring one reason to act to another (IV §4). A person may imagine what her options are in a range of different situations, s1-sn, and prefer one of her options within each of these situations, A in s1, B in s2, C in s3, etc. These options are possible intentions to act that follow from different desires (e.g. the desire to be healthy, the desire to be sociable, the desire to gain wealth, etc.) which would each take precedence in the different situations s1-sn. Furthermore, suppose there is a situation-independent or transcendent perspective from which she prefers intention A (along with its situation and motivating desire) to B to C, etc. When she discovers that all her options A-Q, let us say, are no good because situation s18 obtains, she will not (from the transcendent perspective) be happy that an option as low on her scale as R is the best available, nor is she happy about

---

1078 See Buss, note 23, p.117.

1079 Buss’s analysis seems to require this, although she does not explicitly note this point. Such a transcendent perspective would be one in which preferences are formed in abstraction from many of the facts about our actual situation; judgments about what to do given our motives in this stance are thus not all-things-considered judgments, but only certain-things-considered judgments. And although this is not my strategy, it remains possible for rationalists to construe identification along the lines of the latter (identifications would then be normative for the individual in the same way as judgments in the original position are normative for a society on a Rawlsian model of justice).
the desire which is strongest in s18 (e.g. perhaps the desire to avoid great pain or loss of
desirable physical control). But in acting on R, she is doing what she judges best in the circumstances,
or all things considered. Buss says both her unhappy exerciser and Frankfurt’s unwilling
addict are faced with roughly this sort of problem, and as a result, neither is "helplessly
violated" in performing act R.

But when the point is framed this way, we can see its misunderstanding. The addict’s
being "unwilling," or his alienation of the desire which motivated R (taking the drug, in his
case), is not constituted by his having a preferential hierarchy of the kind just described.
Such a rank-order of preferred situations remains first-order in the strong or qualitative
sense. If identification turned on this kind of an ordering, then the unwilling addict’s
identification with the desire to refrain from drugs would simply be a preference that he laid
aside in favor of his second-best (or in this case, eighteenth best) option, like a person who
has to watch the 18th video on his list because the 17 more-preferred videos are
unfortunately already rented. This is exactly what Frankfurt has rejected in recognizing that
identification involves strong evaluations of motives, as we saw. The addict who does not
identify with his desire for the drug rejects it categorically, despite the fact that he might
prefer, it as his best option in the circumstances. Buss locates a sense in which one may act
optimifically on an option that, in the broadest view of one’s first-order preference rankings,
one is unhappy about: for there is an obvious sense in which I’m not thrilled about watching
the 18th video on my list, but in the circumstances, I rent it quite voluntarily. But the
difference between an option on which I act while being unhappy about it versus an option
on which I act while being happy that it is my optimific option, is clearly not the experienced
phenomenological difference between motives on which an agent acts while alienating them,
and motives with which he identifies.\textsuperscript{1080}

This kind of resistance to or rejection of a motive, (and its opposite) does not seem to arise simply from disliking a situation that lends strength to the motive. In this respect, the unwilling addict is unlike Buss’s putatively unwilling exerciser, who may very well identify with her desire. Her unhappiness about the situation need not involve any rejection of the desire as part of her agency: since this ‘situation’ is constituted by biological facts with which one can neither identify nor alienate—they simply are—whether she internalizes or adopts as ‘her own’ the desire whose strength results from this situation is a separate issue.\textsuperscript{1081} By contrast, the unwilling addict can alienate the ‘situation’ that restricts his options, since his first order motivational set is a fixed constitutive part of that very situation, not just a result of other conditions which fix it. There is, then, a significant difference between the cases Buss compares. In sum, Buss’s reasons for skepticism about the claim that autonomy involves "identification" constituted by some sort of higher-order attitudes actually just point towards the need for a qualitative account of identification, rather than giving us decisive reasons to reject the hierarchical approach to autonomy altogether.

V.5. A Critique of Dennett’s Neo-Behaviorist Account of Identification

V.5.1. Introduction: Dennett’s Cognitive Approach

The previous sections have examined several different ways of explaining identification, including various rationalist or cognitive models, the hermeneutic approach taken by

\textsuperscript{1080}And this seems clear even though I certainly can imagine an addict who is only "unwilling" in Buss’s weak sense. There will be other addicts who are unwilling in a much stronger sense. The fact that epistemically, it may sometimes be difficult to know the difference between them in any given case is not an obstacle to the conceptual clarity of the difference between weak and strong evaluation.

\textsuperscript{1081}As Frankfurt insists, identification with a motivational state is not identification with its cause. I think this principle stands even if some conditions on the origin of identifying states themselves are necessary.
Bransen and Tugendhat, and one of the most promising substantive procedural conceptions of autonomy (Buss’s account). There are other interesting versions of each of these approaches that could be reviewed, but in this brief survey, it will be more useful to turn to the most prominent example of what might be called the pragmatist approach to identification. There are many different forms of pragmatism, of course, but on this approach, most broadly speaking, the notion that some of her motives, dispositions, and action-guiding intentions have special authority for an agent is understood in terms of its heuristic value for explaining or predicting the behavior of agents. What constitutes identification thus depends on when it is useful (in primarily strategic terms) to think of an agent as identifying with or alienating her own motivations.

Daniel Dennett offers an extreme version of this pragmatist approach in his book, *Brainstorms*, which sets out a neo-behaviorist interpretation of personhood in general. Dennett’s account is a good case study for our analysis, however, both because it will allow us to turn back to the wider implications of our findings concerning volitional identification for personhood, and because Dennett’s analysis explicitly tries to accommodate Frankfurt’s “volitional” conditions for personhood within a more general set of conditions that refer only to the intentionality of consciousness and instrumental rationality as basic aspects of human mentality. In other words, Dennett develops a new version of Strawson’s cognitive approach to personhood by incorporating Frankfurt’s second-order volitions into his expanded set of

---

1082 Whether or not in some stronger realist sense, they really have identifications or any other psychological states, as opposed to merely mechanistic states.

1083 Daniel C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Brighton, 1981); chapter 14: “Conditions of Personhood.” This account is now somewhat old, but Dennett’s more recent books have simply developed his account of the intentional stance, of freedom, of consciousness as a system of intentionality, and of the evolution of mind. There is nothing in *The Intentional Stance* (MIT Press, 1987), *Consciousness Explained* (Little, Brown, & Co., 1991), or *Darwin’s Dangerous Idea* that have apparently altered his interpretation of the concept of personhood developed in *Brainstorms*. However there are a few developments in *Elbow Room: The Varieties of Free Will Worth Wanting* (MIT Press, 1984, 1993 pb).
cognitive conditions. But as we will see, Dennett’s attempted reconstruction of Frankfurt’s conditions for personhood fails because volitional identification is irreducible to any set of objective attitudes of the agent or judgments made from the third-personal observer standpoint.

In chapter fourteen of *Brainstorms*, Dennett initially distinguishes between the "metaphysical" notion of a person as "roughly, the notion of an intelligent, conscious, feeling agent" and the "moral" or "forensic" notion of a person who has rights and can be accountable for his actions.\textsuperscript{1084} He believes that persons in the moral sense are a subset of metaphysical persons, and proposes to lay out six conditions, the first three of which will define metaphysical personhood, and all six of which together will capture moral personhood.\textsuperscript{1085} Each of these conditions is a necessary but not sufficient condition for the next, so they are serially ordered and form a set of steps closing in on moral personhood:

"The moral notion of a person and the metaphysical notion of a person are not separate and distinct concepts but just two different and unstable resting points on the same continuum."\textsuperscript{1086}

For Dennett, at the moral end of this series, personhood is constituted by a particularly complex sort of self-consciousness which requires that we are (in a sense) aware of our own intentions, and hence aware of the description that forms a constitutive part of our action.\textsuperscript{1087} This type of consciousness of "the reasons for my actions,"\textsuperscript{1088} as he points out, is a necessary condition for having and acting on (second-order) volitions that are "reflexive" in the sense

\begin{thebibliography}{9}
\bibitem{1084} Ibid, p.268.
\bibitem{1085} Ibid, p.269.
\bibitem{1086} Ibid, p.285.
\bibitem{1087} Ibid, p.283.
\bibitem{1088} Ibid.
\end{thebibliography}
of being about *my own* (first-order) desires. Thus Dennett believes what is significant in Frankfurt’s analysis can be captured by his own analysis, once we have defined the right sort of reflexive consciousness of one’s intentions: it is this which gives a sense of our *self*.

Dennett’s first condition, which is supposed to establish personhood in the barest ‘metaphysical’ sense, is a wide mental condition which will allow further differentiation by other conditions. This condition is that the entity exhibit *rationality* in the purely instrumental sense of strategic planning in pursuit of ends in its outward behavior—a theme which Dennett *claims* to find in authorities such as Kant, Rawls, Aristotle, and Hintikka. The next condition is that "persons are beings to which states of consciousness are attributed, or to which psychological or mental or *intentional predicates*, are ascribed." This condition is incorporated in Strawson’s notion of the person as an entity to which both physical predicates and intentional predicates of conscious states apply. The further conditions all specify more complex forms of intentionality: (3) the entity is able to be interpreted according to the "intentional stance"; (4) the entity is capable of interpreting *others* via the "intentional stance," or via attributing intentional states to them; (5) the entity is capable of verbal communication with others. Finally, Dennett tries to show that the phenomena which Frankfurt brought to light against Strawson can be expressed in terms of a final further condition of the *same general kind*, which will be the sufficient condition for personhood in the fullest *moral* sense:

The *sixth* theme is that persons are distinguishable from other entities by being *conscious* in some special way: there is a way in which *we* are conscious in which

---

1089 Ibid, p.284.
1090 Ibid, p.269.
1091 Ibid, p.269.
1092 Ibid, p.270.
1093 Thus making it a "second-order intentional system" in Dennett’s sense.
Dennett begins from the supposition that volitional identification is simply a mode of self-consciousness, a way in which agents cognize themselves, or take reflective attitudes towards themselves, or intend their own intentional states. As we will see, however, this assumption leads Dennett in the wrong direction. As we saw in discussing developments in Frankfurt’s theory of identification, he argues that such self-regarding attitudes can always be both "internal" and "external" to the person’s identity: hence "an infinite regress will be generated by any attempt to account for internality or externality in terms of attitudes" rather than decisions.1095

V.5.2. The Intentional Stance: Pragmatism, Neo-Behaviorism, Reductionism

Let us take a closer look at conditions (3) - (6) of Dennett’s analysis. The third condition is derived from Dennett’s account of intentional systems:

An intentional system is a system whose behavior can be (at least sometimes) explained and predicted by relying on ascriptions to the system of beliefs and desires (and other intentionally characterized features—what I will call intentions here, meaning to include hopes, fears, intentions, perceptions, expectations, etc.).1096

For example, we may regard a chess computer or even a tree as an intentional system if

---

1094 Dennett, p.271.
1095 Frankfurt, "Identification and Externality," p.66. Dennett is also wrong to describe Sartre’s conception of the "for-itself" as an instance of his sixth condition. For the self-consciousness which concerns Sartre is that of the pre-reflective cogito, which is precisely not a reflective or second-iteration intentional state about intentional states. Moreover, in Being and Nothingness, the type of "presence to self" which characterizes the for-itself is manifested in volitional freedom for its original project. The nothingness which makes the for-itself ‘be what it is not and not be what it is’ is thus in two ways more than a reflective self-regarding attitude.
employing this fiction is helpful to us.\textsuperscript{1097} It is important that what counts as an intentional system in this sense is not determined by its "really having" the intentional states Dennett mentions here: he is trying both to avoid having to define what counts as "really having a belief," for example, and to avoid assuming that anything really does have such states in our folk-psychological sense of them.\textsuperscript{1098} Dennett thus intends his approach to leave open the possibility that entities satisfying his third condition (and as it will turn out, the other conditions as well) are merely complex mechanisms whose states would not, in absolute metaphysical terms, need to be described in terms of intentional or psychological states. In principle, this description may be dispensible.

Being an intentional system, then, is not a purely metaphysical property of an entity, but rather an observer-relative property, defined by the value of taking an "intentional stance" towards it. For A to take an (first-order) "intentional stance" towards B in Dennett’s account is for A to regard B as a (first-order) intentional system, and for A’s so regarding B to be rational for A in a strategic sense, since this is the most feasible and powerful way for A to explain and predict B’s behavior-patterns.\textsuperscript{1099} The first two conditions are then reinterpreted in terms of this notion of "intentional stance." Thus a being is "rational" in the sense required by the first condition if and only if the entity acts as if it has purposes, displays activity interpretable as goal-seeking etc.\textsuperscript{1100} So this condition now amounts to its being a good strategy to attribute purely strategic intentionality to a system. Thus, as McCall points out, Dennett’s "conditions of personhood describe not a metaphysical but a social

\textsuperscript{1097}Ibid, p.272.
\textsuperscript{1098}Ibid, p.272-73.
\textsuperscript{1099}Thus even if in principle A could instead explain and predict B’s behavior in terms of machine language algorithms or particle physics, this might be mathematically impractical or too cumbersome to be viable.
\textsuperscript{1100}Dennett, p.271-272. For example, the tree is an intentional system because it can be regarded as wanting to "spread his limbs" etc.
This clarifies the sense in which Dennett’s approach to personhood is not simply social, but pragmatist, neo-behaviorist, and potentially reductionist.\textsuperscript{1102} His third condition says more than that what attitudes we take towards an entity are "somehow and to some extent constitutive of its being a person."\textsuperscript{1103} Any social conception of personhood whatsoever agrees with this general thesis that an entity’s being a person cannot be a fact that is independent of that entities’s social relations to other similarly situated entities. But many social conceptions would not share Dennett’s anti-realist construal of what it is simply to have intentional states.\textsuperscript{1104}

Dennett’s account of personhood is not merely social but \textit{pragmatist} in character, then, because in his view, persons are a particularly sophisticated sort of intentional system, where this means only that it is pragmatic to \textit{regard them as if} they were sophisticated in a particular sort of way (which conditions 4-6 will spell out). This account is also \textit{neo-behaviorist} (as many pragmatist conceptions are not), because its being useful or pragmatic to take an intentional stance with regard to an entity is cashed out by Dennett in purely self-serving instrumental terms. As Dennett says, regarding the other entity as an intentional system is "simply the most efficient way of making sense of, controlling, predicting, and explaining the behavior" of the other entity.\textsuperscript{1105} The model is that of game-theoretic rationality: if it is a "good...strategy" to regard an entity as an intentional system in predicting


\textsuperscript{1102}Dennett’s term for his approach—\textit{instrumentalism}—obscures these distinctions (see \textit{The Intentional Stance}, pp.69-81).

\textsuperscript{1103}Dennett, \textit{Brainstorms}, p.270.

\textsuperscript{1104}For many (perhaps most) social conceptions of personhood allow that an entity’s having psychological states which can only be analyzed in terms of intentional predicates \textit{can} be an independent fact about that entity alone, although the additional features of its psychological life needed to make it a person (such as recognition as an agent by others, etc.) are not similarly independent of interpersonal relations.

\textsuperscript{1105}Ibid, p.272.
its behavior, or if this tactic will contribute to our success in realizing other ends, then it is an intentional system. Finally, this account is reductionist since it suggests that personhood, like the intentional properties in terms of which it will be defined, may only be a fiction we continue to find useful; even if it is indispensable in practice, it need not imply anything metaphysically deep about human beings or other entities to whom it is applied.

There are several immediate objections that can be pressed against Dennett’s account at this point, although I will only mention them in passing. First, there is a fairly obvious regress problem in his account of intentionality that undermines his reductionist goal: for A to take an intentional stance towards B itself looks like an intentional state of A’s; hence mentality or intentional states cannot be cashed out in terms of intentional stances all the way down. If we interpret mentality in terms of intentional stances, we will always be left with an assumed but unexplained surplus of intentional states, namely those constituting the intentional stances whose ‘existence’ is not itself a result of other intentional stances.

Second, we ordinarily make a distinction between merely regarding something as if it were a person for some external purpose (e.g. in the fiction of treating a bank corporation as a person to simplify legal proceedings) and regarding it as being a person. But on Dennett’s account, there is no way of distinguishing these attitudes: reality and metaphor are collapsed. Hence it is not surprising that corporate or institutional entities such as the IRS turn out literally to be persons on his account (a result which some may take immediately as a reductio of his analysis).

Third, unlike Strawson, Searle, and others, it is not obvious on Dennett’s account that the intentionality attributed to entities connects with sentience or consciousness in the ordinary sense at all. Dennett can ignore the immanent reflexivity or pre-reflective self-awareness

\[1106\] Ibid, p.271.
necessary actually to experience intentional states, because his second condition only requires that we find it useful to act as if a system had strategic rationality, without even considering whether, from its own first-person perspective (if there even is such a perspective) there is something it is like for it to be acting purposefully. From the external point of view of Dennett’s so-called "heterophenomenology," there is little difference in heuristic value between regarding something as having conscious intentional states vs. regarding it as simply having intentional states: as Dennett says, "it is far from clear that all or even any of the beliefs and other intentions of an intentional system need be represented ‘within’ the system in any way for us to get a purchase on predicting its behavior by ascribing such intentions to it."\textsuperscript{1107} But this difference is hardly irrelevant to personhood. The Strawsonian condition of personhood which Dennett’s second condition purports to reconstruct requires that persons have consciousness, not merely that within them, states occur that could be construed as about something.\textsuperscript{1108}

Similar objections can be raised about Dennett’s fourth and fifth conditions, which refine the picture of the type of intentional system involved in personhood. The fourth condition for personhood is a capacity for reciprocity like that normally associated with interpersonal activity. But Dennett defines this reciprocity in terms of "second-order intentional systems," which are interpreted in the same pragmatic fashion: an entity has this capacity for reciprocity if and only if, from a neutral third-person observer position, it would be useful to regard it as a second-order intentional system, that is, one that takes first-order intentional stances towards others. And this is not unique to human beings: it may be useful to regard a dog or a bird as a second-order intentional system if their behavior is explained by assuming

\textsuperscript{1107}Ibid, p.277.

\textsuperscript{1108}This is related to the difference between semantic vs. syntactic intentionality. See Dennett’s discussion of his debate with Searle in The Intentional Stance, p.70-71, and chapters 8-9.
they sometimes intend to induce false beliefs in other entities. In other words, for X to be a second-order intentional system is for it to be pragmatic to view X as if X both has intentional states and strategic rationality, and some of X's intentions attribute intentional states and strategic rationality to others as a means to its own ends. Thus, although he does not emphasize this sufficiently, in Dennett's terms, second-order intentional systems are not necessarily those which actually take intentional stances towards other systems, but only those to which, "for heuristic and pragmatic reasons," we ascribe second-order beliefs and intentions about other systems's intentions. This pragmatically attributed reciprocity of strategic motivations is rather different than the reciprocity normally accorded to persons: for example, reciprocity in Nagel's sense, i.e. as my understanding that another entity really has a first-person perspective irreducible to any of my third-person descriptions of it, and even more clearly irreducible to what might be useful to think about from the narrow point of view of my own interests. Reciprocity on Dennett's account is that of fictive or imaginary Hobbesian agents.

In interpreting his fifth condition, Dennett borrows from Grice's reductionist explanation of the capacity for verbal communication of "non-natural" meaning. Grice's theory explains such meaning in terms of strategic third-order intentions: to recognize an interlocutor (V) as engaging in verbal communication, I must attribute to V the intention that I come to recognize that V intends to produce a response in me (that serves his purposes).

---

1109Ibid, p.275-6. Dennett cites the example of a low-nesting bird leaving the nest and dragging its wing as if broken to lure the fox away.

1109Dennett, Brainstorms, p.273.

1111Ibid, p.277-8. There is no need to demonstrate the inadequacy of Grice's theory (and thus Dennett's) at this level, for this has been done with devastating thoroughness by Jürgen Habermas in his "Toward a Critique of the Theory of Meaning" in Postmetaphysical Thinking (MIT Press, 1992). As Habermas points out, Strawson already demonstrated that Grice's theory allows for cases in which S can only be successful in affecting his interlocutor H as intended, if "the intention that H is supposed to take for the intention of S is not identical with the strategic background intention that S is actually pursuing" (p.65). Since this case leads to an (continued...)
Similarly, in order to show how his first five conditions (up to the capacity for verbal encounter) "might play a role in ethics," Dennett interprets Rawls's derivation of principles of justice from the original position in terms of strategic rational choice theory. On this account, justice as a norm for interpersonal relations is reduced to the principle that a third party rationally calculates that the parties behind the Veil (the constraint representing 'the reasonable') would adopt as most strategic.\footnote{\pageref{fn:1112}}

These problems with Dennett's first five conditions, significant as they are, are not the main focus of my attention here. However, they prepare us to see the fatal difficulty in Dennett's sixth condition, which attempts a pragmatist reconstruction of volitional identification.

\section*{V.5.3 Dennett's Reflective Self-Manipulator Model of Identification}

Dennett believes he can capture Frankfurt's notion of second-order volitions by extending his own notion of a third-order intentional system (i.e. verbal communicator) to the self-referential case. This extension makes sense, Dennett claims, because having and acting on second-order volitions requires "adopting towards oneself the stance not simply of communicator but of Anscombian reason-asker and persuader."\footnote{\pageref{fn:1113}} This means regarding oneself as, or acting as if one is, an Anscombian agent, which Dennett explains as follows.

\begin{itemize}
  \item [\footnote{\pageref{fn:1111}}] (\ldots continued)\n\end{itemize}

\begin{itemize}
  \item infinite regression, as Strawson showed, the fact that such deceptive verbal communication is possible leads to a \textit{reductio} of Grice's theory: the exercise of \textit{strategic rationality} through verbal communication cannot account for the fundamental level of shared meaning present in communicative action, but can only be parasitic on this trans-strategic type of meaning. This argument is an essential part of Habermas's demonstration that \textit{communicative} or "verbal" non-natural meaning is irreducible to any "semantically analyzable knowledge of success conditions" (p.67).
  \item Of course, it is precisely the danger that Rawls's theory may be recast in these decision-theoretic terms that leads Habermas to object to the mechanism of the original position as adequately capturing the sense of justice. See Habermas's \textit{Discourse Ethics} and his review of Rawls's \textit{Political Liberalism}.
  \item Dennett, \textit{Brainstorms}, p.284
\end{itemize}
An Anscombian agent is a being that is aware of its own intentions—and hence could give a
description of its acts—in the same way as one is (allegedly) aware of the third-order
intentions under which another entity is acting when one is aware of that entity as trying to
communicate something.\textsuperscript{1114} Anscombian agency is the self-referential case of verbal
communication: the only difference is that in the Anscombian variant, the ‘person’ with
whom I am communicating \textit{happens} to be myself. But I communicate with him just as I
would anyone else: I have no privileged access, but must interpret what he intends me to
recognize from his behavior.

Thus being an Anscombian agent in this sense is really only a \textit{specific case} of the fifth
condition, involving no essential difference. For Anscombian agency, \textit{as Dennett reads
it},\textsuperscript{1115} simply means treating oneself heuristically \textit{as if} one were a third-order intentional
system (just as one treats one’s interlocutors). This is what it means to \textit{treat oneself} as
having intentions under which one acts, including (paradigmatically) intentions to
communicate or express one’s intentions to others. Such is Dennett’s third-personal
translation of what it is to be aware that one has a \textit{first-order will} in Frankfurt’s sense,\textsuperscript{1116} i.e.
a motive operative in one’s action.\textsuperscript{1117}

The sixth and final condition for personhood is that an entity can regard herself and
others as Anscombian agents (who regard themselves as third-order intentional systems).

\textsuperscript{1114}Ibid, p.283.

\textsuperscript{1115}For I do not suppose that Anscombe would actually agree with Dennett’s interpretation.

\textsuperscript{1116}Note that the differences in Frankfurt’s "orders" and Dennett’s "orders" are significant. Frankfurt’s
orders were intended to pick out levels of \textit{identification} or agent-authority (though as orders of desire they
failed at this task); Dennett’s orders only indicate \textit{iterations} in pragmatic ‘as if’ interpretations, i.e. orders of
\textit{explanatory behavior}. Thus it is no surprise that these orders diverge: as Dennett fails to realize, they are in
fact incommensurable.

\textsuperscript{1117}One interesting implication of this is that Dennett has implicitly admitted that Grice’s definition of verbal
communication is inadequate: when I regard my interlocutor as communicating with me, I do not simply treat
her as if she were a third-order intentional system; I regard her already as an Anscombian agent, i.e. \textit{as if she
regarded herself} as a third-order intentional system, and thus had an interpretation of her own motives for
acting. Thus Dennett’s fifth condition really expands into his sixth. The two are not formally distinguishable.
Dennett then claims that Frankfurt’s requirement that persons be capable of higher-order volitions turns out to be equivalent to this requirement: "Frankfurt claims that persons are the subclass of intentional systems capable of what he calls ‘second-order volitions.’”  

This does not mean that persons are simply second-order intentional systems, however; rather:

...the "reflective self-evaluation" which Frankfurt speaks of is, and must be, genuine self-consciousness, which is achieved only by adopting towards oneself the stance not simply of communicator but of Anscombian reason-asker and persuader....Acting on a second-order desire, doing something to bring it about that one acquires a first-order desire, is acting upon oneself just as one would act upon another person: one schools oneself, one offers oneself persuasions, arguments, threats, bribes, in hopes of inducing oneself to acquire the first-order desire. One’s stance towards oneself is the same as one’s stance toward and access to another.

Thus on Dennett’s account, identification with motives or dispositions we would bring about in ourselves is equivalent to adopting a strategic attitude towards our reasons for acting, or adopting a fourth-order intentional stance towards ourselves.

Our foregoing analysis of identification in Chapter IV allows us to see how radically mistaken this account is. First and foremost, it is clear that the kind of relation among desires which Dennett describes is the same kind as found in the relation among second- and first-iteration desires in a complex preference to maximize one’s desire-satisfaction. But as we have seen, such mere iteration has nothing essential to do with expressing the agent’s authority. An agent might have a motive$^2$ to cajole and manipulate himself in the way Dennett suggests, and yet not identify with this motive$^2$. For example, a homosexual man could be trying to argue or coerce himself into feeling sexual attraction towards women (just as a third-party might) in order to live up to his parents’ expectations, yet inwardly alienate

---

1118 Ibid, p.283.
this entire effort and identify instead with the desire to ‘come out.’ Dennett’s model fails to satisfy the Qualitative Condition in the same way as Frankfurt’s initial model failed to.

Second, Frankfurt’s theory of what it is to care about something gives us another obvious reason to think that my second-order volition is not an attempt to manipulate myself (for this is Dennett’s understanding of self-communication) in the same way as Dennett assumes I do when I am trying to convince another person (someone I regard as an Anscombian agent) to change her intentions. For one of the things I can try to convince her to do is to care about things that ought to be important to her, and even to care about the persistence and stability which caring itself brings. But having a steady and enduring second-order volition about my own first-order desires, just is to care about the project or purpose to which those desires relate. In caring, then, I am already related to myself in a way that makes redundant the type of persuasion involved in coaxing another to change her intentions. Even if I am trying to develop motives and dispositions I presently lack, it is the resolute projection of this end, and not the conversation I might or might not have with myself about this second-order project, that constitutes my identifying with those motives.1120

Third, and perhaps most disturbingly, Dennett’s pragmatic reconstruction of Frankfurt’s account is not even true to Frankfurt’s initial model, because it reduces the essentially intrapersonal character of Frankfurt’s volitional relation to a set of attitudes towards myself as an object rather than as a subject. In a Dennettian second-order volition, I simply take the same kind of manipulative attitude towards myself as I would take towards another person whom I regard as an Anscombian agent. This makes it impossible to explain how my second-order volition can identify me with first-order intentions I already regard not simply

---

1120 Thus even if Dennett’s false account of communication as strategic action is replaced by the correct Habermasian theory of shared ideal conceptions of validity in cooperative action, the difference between intrapersonal and interpersonal communication remains significant.
as someone’s, but as mine (or, following my earlier description, as being at least minimally identified with me in the first-person sense). Yet it is not enough that I have intentions about some other intentional states in abstract: it is essential to identification, even as Frankfurt initially described it, that it conveys authority to states that are already recognized as my own in one important sense.

In fact, Dennett’s account is useful as a foil because it lets us see the importance of this feature of identification more clearly: the identifying agent has to be able to recognize an immediate identity between "herself" as identifier and "herself" as the agent who acts on the first-order motivational states towards which she is projecting her will in identification. This immediate relation is not a product of the higher-order activity at issue here, but rather its presupposition, and so in identification, we are not originally taking ourselves merely as objects, or regarding ourselves from a removed reflective stance. Thus an analysis of identification cannot dispense with its irreducibly intrasubjective character without being an eliminitivist analysis that diagnoses these volitional phenomena as illusory (which is eo ipso implausible).

By contrast, Dennett’s pragmatist reconstruction suggests that personhood is constituted by a kind of strategic effort in interpreting an alien being called ‘one’s self.’ Dennett’s construal of the final criterion for personhood ignores that even Anscombian agency, or consciousness of one’s own intentions or the description under which one acts — the consciousness that enables one to answer questions such as ‘What do you think you are doing?’ — already implicitly includes a kind of identification which can in principle never be explained by conscious attitudes alone. This can be seen whether we start with a weaker or stronger conception of "action.”

Suppose we assume that action consists in no more than performances that involve
reasons, such as desires and preferences, which involve descriptions of the action including its end; thus ‘actions’ on this account involve the "intention" of "ends" in the suitably defined sense. Frankfurt’s argument is designed precisely to show that even if action in this sense involves such a reflexive consciousness of the intentions operative in one’s act, such consciousness is by itself totally insufficient for personality and moral responsibility. The wanton may be fully aware of the first-order desire on which he acts; he may be able to come up with a precise description which captures the intention that motivated his performance. Thus, in performing the action, he may have in mind a first-order purpose or end and a means to pursue that end (a first-order intention or action-maxim): for example, ‘I will take some crack-cocaine in order to get high.’ This constitutes a full intention in Anscombe’s sense, or an action-maxim in Kant’s sense. And yet, Frankfurt argues, this reflexive practical consciousness tells us nothing about whether its owner identifies with it as a person. This addict may be entirely wanton or he might be unwilling: in either case, he is not authentically identified with his Anscombian intention or first-order action-maxim. Authentic identification requires a condition of the higher-order will, for which this mere reflexive consciousness of one’s first-order intention does not suffice.

The force Frankfurt’s insight has against Dennett’s account can be made clear with a hypothetical example. Imagine an animal (or a computer, if you like) with highly advanced cognitive powers and abilities to interpret and make fine discriminations about its own mental states. Let us even add for good measure (since this is not the same thing) that this animal (or computer) has reflexive consciousness of its own responses to stimuli, and also awareness of its ends in actions like (say) taking the dog out for a walk so it can get some exercise. It thus has first-order intentions in the proper sense, since it is conscious of the motives on which it acts —thus it acts ‘under an intentional description.’ It can "act" in this
sense and even regard itself as an Anscombian agent. Yet Frankfurt’s analysis points to the reason why, for all its abilities, this animal (or device) is still not a person: its ability to make sense of its actions, to interpret their meanings, still gives it no power to will anything with respect to the motives behind its intentions. It can think about these motives and the intentions formed on their basis, but it remains essentially neutral towards the fact that its intentions are grounded in such-and-such a motive. It cannot care about the ultimate meaning of its actions. Without the power to identify with the motives on which it acts, it can never be anything but a wanton. We can label it WHA: the ‘wanton higher animal.’

WHA has all the powers of consciousness and advanced cognitive features that are essential to human beings, but it is not a person. Its conscious awareness of the meanings ascribable to its actions is simply a cognitive extension of the lower affective reflexivity in all consciousness: it lacks the irreducible volitional condition essential for identification, moral responsibility for its first-order character, personhood in the ‘forensic’ sense.

On the other hand, we might define action in a stricter, existential sense as including identification. I do not act at all in this existential sense if I am not authentically identifying with pursuing the end or purpose in the description under which I perform my action: in Kantian terms, I must have incorporated an intention picked out by this description as the maxim of my action. To act on a first-order desire, in this specialized existential sense, requires identification with its intended goal, and such authentic identification with the end of my action may require reflexive consciousness. But in that case, the constitution of my performance as an action still depends directly on my identification with its motive, and only indirectly on the self-reflecting consciousness. This is the point Dennett entirely misses.

If we defined action and intention in this existential sense, once again it would be
obvious that WHA lacks the capacity to act in this way, and thus lacks the full moral responsibility necessary for personhood. WHA only has a *third-party* consciousness of the goal which its "intentional system" (in Dennett’s sense) is seeking. This can even be a "third-order intention system" as Dennett defines it. WHA is aware of its own *response* to verbal communication in the same way: but this is precisely to be aware of the response just as *some* entity’s, rather than as *its own* in our existential sense. For example, WHA knows that its performance of speaking-motions can be understood under the description of ‘response to another’s intention to verbally communicate something.’ But this is still nothing more than a disinterested, third-party observation WHA makes about its behaviors: it does not entail any *personal commitment* of a self. WHA does not existentially *act* with the existential *intention of responding*; rather, it only knows that the motions of its vocal cords and the desires which preceded them could be interpreted that way by a neutral observer, if he had some strategic purpose for engaging in such interpretation. ⁴⁵⁹ But in principle, WHA is entirely neutral about the possibility of this interpretation: it is not *volitionally relevant* for WHA.

In sum, Dennett’s reflexive consciousness of intentions involves cognitive complexity but no essentially *first-personal authentic identification* anywhere along the line. And without this volitional condition, it is impossible that persons really project the intention under which they know their act could be described: for them to be *identified* with this intention, they must will to act on this intention, in order that *they themselves* acquire one sort of ‘outward character’ rather than another.

---

⁴⁵⁹ A person may also *act* on only one of the intentions he knows others could *attribute* to him as a result. Thus—pace Davidson *et. al.*—making a certain performance with the conception of it under a given description, is still *insufficient* for committing the *action* that involves that description. This point is akin to the distinction between foreseeing and intending consequences of one’s action.
V.6. Conclusion: Volitional Identification and Personhood

V.6.1. The ‘Mentalist’ Tradition Behind’s Dennett’s Pragmatist Interpretation

In the Introduction to this dissertation, I noted that an existentialist approach to personhood is constituted in part by an opposition to what I called the mentalist approach, which assumes that the distinguishing features of personhood have only to do with our consciousness and our rationality. As argued in the Introduction, this is the assumption that drives all classical philosophical anthropology: since mind or ‘rational soul’ is the distinguishing mark of humanity as a natural kind, conditions capturing what it is to be conscious and rational will jointly capture what it is to be a person.

John Locke. This approach to personhood has a background long predating Dennett’s own neo-behaviorist version of pragmatism. In particular, Dennett owes a special debt to John Locke, who held that what makes (or would make) a person is a special sort of consciousness, traditionally involving memory or some other sort of reflexive consciousness or unifying function as the basis of the capacity for reidentification of oneself. This view has continued to be popular in the philosophy of mind ever since Locke. To cite just one example, Richard Wollheim acknowledges Locke’s emphasis on the continuity of memory in his own attempt to explain "the highly significant fact about our identity that it is such, or is so grounded, that ..it does give rise to a sense of itself" by way of a particular kind of

---

1122 See Locke’s An Essay Concerning Human Understanding, ch.XXVII. To his credit, Locke was the first philosopher to explicitly distinguish being human in the "species" sense from being a person (see §10-13 especially). But although he distinguished the question of personality as a "forensic" one (§26) directly related to the conditions sufficient for moral responsibility for one’s actions, the fact that Locke finds sufficient conditions for responsibility in continuity of consciousness shows that his recasting of the question did not lead to any radical break with the mentalist tradition; quite the contrary, he adhered strictly to the deepest classical assumption that it is some aspect of mentality which accounts for forensic personhood. Thus, when he abandoned the substance analysis of the mental conditions for personhood, he was forced to look to relations and functions of consciousness for an explanation—like modern mentalists today.
Dennett’s debt to Locke is also apparent when he attempts to interpret the "special sort of consciousness" appropriate to moral agents as an extension of the linguistic ability to recognize communicators’ intentions.

**Clark and Parfit on corporate personhood.** Philosophers of mind guided by Dennett’s model always accept the mentalist presupposition as their ultimate starting point. To take a recent example, Austen Clark has argued that "every condition that [Daniel] Dennett specifies for adopting the intentional stance can be satisfied as well by corporate entities," and hence that entities such as governments can be regarded as having "beliefs and desires." Thus it is just as true for corporations as for normal human beings that attributing beliefs and desires to them yields "predictions which were not antecedently available," thus satisfying Dennett’s criteria for the intentional stance. This parity of reason argument is already weak because it depends on the suspect assumption that the only sense in which anything ‘has’ beliefs and desires is the reductive pragmatic one: i.e. it is useful to regard these entities as having beliefs and desires. But from this already-weak conclusion, Clark makes what is in fact a further inference, although he does not distinguish it as such: since corporations have beliefs and desires in the same sense as we do, we are justified in "treating a corporate entity as a person."

By this, Clark does not mean that we should regard corporations as persons in a derivative sense, as has been common since Hobbes. For he goes on to argue that persons in

---


1124 Dennett, Brainstorms, p.271.


1126 Ibid, p.409.

the ‘primary’ sense never were any less ‘corporate’ than, say, Simon & Schuster. Clark claims that connectionist theories of mind provide an explanation for why corporate entities should turn out to satisfy Dennett’s criteria: these theories of mind suggest that "...in fact persons are corporate entities." The distinction between persons in the traditionally primary sense and institutional entities such as the Bank of America rested on a false (i.e. non-pragmatist) basis, so there is no reason to regard the latter as only derivatively a person in the legal or forensic sense. Clark’s warrant for this shocking claim comes from "homuncessional" cognitive psychology:

Homuncularism has empirical roots; the claim that a person is a corporate entity is an empirical claim. Indeed, it draws its support from ongoing research programs in cognitive science. Beneath the disagreements between (say) Jerry Fodor and the connectionists, there is at least one claim on which all parties agree: the mind is a massively parallel system...

There can be no doubt that Clark has accurately located the point of convergence between a number of reductionist approaches to personhood, including Dennett’s analysis of attributability of instrumental intentionality as the essence of personhood, Fodor’s modular model, and Derek Parfit’s argument that personal identity is reducible to an ensemble of more particular facts. But there is a deeper agreement common to them all which Clark does not highlight: namely, the unquestioned yet fundamental assumption that a person is identifiable with their mind, and in particular, with their capacity for reasoning and intentional consciousness, whether real or merely attributed (call this the ‘mentalist presupposition’).

---

1129Ibid, p.412. It does seem, however, that Clark’s parity of reason argument depends particularly on Dennett’s pragmatist reduction of beliefs and desires (which extends the behaviorist approach to allow for the ‘attributed’ existence of such states). On Fodor’s and Parfit’s accounts, it seems unlikely that any human corporation could approach the degree of parallel integration necessary for the kind of mentality constitutive of persons. So the distinction between persons in the primary and derivative sense would still hold.
All the reductionist conceptions of personhood Clark considers, whether merely explanatory (as in Dennett’s case) or ontological (as in Parfit’s case), depend not incidentally but essentially on the mentalist presupposition, because they take evidence that the mind or consciousness can be understood as a ‘corporate’ ensemble, which functions in ways that can be described as ‘end-seeking,’ as immediately equivalent to evidence that a person is nothing more than such a functional organism. The crucial role played by the mentalist presupposition in these approaches is evinced by one of Clark’s own remarks. After quoting Hume’s comparison of the soul to a "republic or commonwealth," Clark notes that "Contemporary functionalism holds that this is no mere analogy: people (or their minds anyway) are corporate entities."\textsuperscript{1130} The entire story of the contemporary analysis of personhood in analytic philosophy of mind is contained in this beautiful finesse: persons, or their minds, anyway... As long as the mentalist presupposition holds sway, this inference from mind to person is taken for granted.\textsuperscript{1131}

Clark is therefore absolutely wrong when he says that "the claim that a person is a corporate entity is an empirical claim." Even the weaker claim that the mind’s corporateness is empirically demonstrable would have to face a number of counterarguments for the irreducibility of typical features of mentality, such as the ‘feel’ of qualia and the immediate reflexive awareness of consciousness. To take just one example out of many, there appears to be no non-circular explanation for how our awareness of non-actual possibilities of any sort can be constituted by some set of actual relations among states of different parts of our brain. The openness to modality which is essential to noesis for Plato

\textsuperscript{1130}Ibid, p.410.

\textsuperscript{1131}One obvious response to Clark’s argument might be that if the mentalist presupposition leads to so absurd a result as Abraham Lincoln’s being no more a primarily a ‘person’ than General Motors, we should immediately and thankfully take it as a reductio of the mentalist presupposition.
and to "thinking" (denken) for Heidegger is very far from having been explained empirically. But even if we hypothetically granted Clark that every feature of consciousness and human rationality can be ‘naturalized’ or given a complete account in physicalist terms, it will not follow that personhood gets reduced with them. The reduction of personhood in general to any form of corporate existence on this basis depends on the mentalist presupposition, which in principle cannot be established by any direct empirical evidence. This presupposition involves a fundamental albeit largely unexamined interpretation of personhood which determines the relevance of empirical evidence about features of the mind: it is a synthetic apriori presupposition derived from an unsupportable classical metaphysics.

**Emotivist Reactions to Rationalism.** Although I have focused on mentalist assumptions in philosophy of mind as inspired by advances in artificial intelligence and neurobiology, in moral psychology, contemporary thinkers have also looked to passions and emotions as the basis for conceptions of value which make us human. The ‘pre-ethical’ conceptions of agency which emotivism provided for utilitarian ethics give us our clearest twentieth century examples of this approach to personhood. The emotivist approach broadly construed, which takes our capacity for affective states, such as needs and desires, to be the only possible motivational ground for our use of reason (which can then never exceed a purely intrumental function), also depends on mentalist presuppositions. Emotivism sees personhood as a result of possessing certain sorts of affective states of consciousness, and it does not distinguish volitional states in general from non-cognitive passions, desires, or

---

1132 This is the basis for what I call a modal argument for libertarian freedom, which does not depend on any inferences from the conditions of moral responsibility.
tendencies. Since its proponents, unlike their scholastic forebearers, tend to be compatibilists of various sorts, the will becomes for them a mere repository of passions, which are distinguishable from the more ‘brutal’ instincts only by their ability to be motivate rational calculations of long-term consequences and strategies.

Note that denying that first-personal motivation can arise directly from impartial reasoning does not by itself entail an emotivist conception of personhood. Emotivism requires the additional step of inferring, from the limited motivational powers of impartial reflection, that the personal involvement or "identification" necessary for moral responsibility for one’s actions arises strictly from acting on personal preferences, emotions, and desires. Moreover, an emotivist must conclude that continuity of such non-cognitive states is the sole sufficient condition for personal identity over time.

Bernard Williams’s is a good example of someone who denies the motivational power of ‘pure practical reason’ without falling into a purely Humean, emotivist view of the self. In general, Williams defends first-person identity as irreducible: for example, in through an ingeniously constructed antinomy in "The Self and the Future," Williams shows that the notion of my presence in the future seems not to depend necessarily on either continuity of body or continuity of mental features such as memories, beliefs, and emotions. Rather, the unity and continuity of first-personal identity appears to be irreducible. Nevertheless, Williams regards reason by itself as too abstract a basis for motivations which are essentially relativized to the particular person: it is for this reason that Kantian formalism as a

\[1133\] For example, on this approach, love in the agape sense, which is a volitional state on my account, can be distinguished from simple attraction or the infatuation of eros only by their ability to be motivate rational calculations of long-term consequences and strategies.


justification for impartial motives provides "ultimately too slim a sense in which any projects are mine at all."\textsuperscript{1136} Clearly, however, this does not mean that Williams holds an emotivist conception of personhood. Rather, his critique of Kant relates to the question of whether persons, whose first-person identity seems to be irreducible to their purely mental features, can be motivated by universalistic reason.\textsuperscript{1137}

The emotivist conception of personhood thus largely begins with Humean’s approach to reason and the passions, and fits with Locke’s theory of the self: we do not need to identify the self with a will that is distinguishable in any way from the series of its affective and emotional states, for "the passions" are not unified in an individual human being by any ‘self’ with more content than mere continuity of consciousness.

\textbf{V.6.2. Why the Radical Implications of Frankfurt’s Analysis Were Missed}

Along with the mentalist presupposition about personhood deriving from classical philosophical anthropology goes what I will call the \emph{extendability assumption}: since the entire structure and essential conditions of personhood are defined in terms of consciousness and rationality, the terms in which sentience, self-awareness, theoretical reason and practical reason are understood can always be \emph{extended} to any new features found phenomenologically to be indispensable for personhood.

Dennett’s analysis of Frankfurtian identification illustrates the prevalence of this ‘extendability assumption.’ Rather than considering that Frankfurt might have discovered


\textsuperscript{1137}It may be said against Williams, however, that he systematically conflates the \textit{third-person} ‘objectivity’ of scientific judgment (‘objectivity’ in Nagel’s sense) with the “impartiality” of the Kantian "moral point of view" (e.g. see his "Persons, character, and morality," p.2). Williams does not seem to realize that "a rational application of an impartial principle"—as he describes the operation of pure practical reason—may nevertheless have an entirely \textit{first-personal} significance for the individual will. Necessity vs. contingency in moral judgment and third-person vs. first-person significance are \textit{different distinctions}. 
important volitional features of personhood that a self-conscious rational being could still lack, Dennett assumed that "second-order volition" or what it represented could just be assimilated into the broader mentalist picture by explaining it in terms of a certain rational ascription of cognitive features to oneself. To be sure, Dennett thought that Frankfurt had shown that Strawson’s cognitive conditions were not sufficient for personhood: but like other philosophers of mind, he missed the implication in Frankfurt’s discovery that no set of cognitive conditions of that kind would ever be adequate to explain identification, since identification requires that a person give to (or recognize in) her actions and motives a meaning which essentially refers to her own self (even though her assessment of it may not be infallible).  

Frankfurt’s contribution was taken, then, simply as implying the need to broaden the set of conditions relevant for understanding the cognitive and intentional conditions characteristic of persons—not as evidence that there are volitional features of personhood irreducible even to reflexive forms of practical rationality, desire, and consciousness of one’s states. Thus despite Frankfurt, in contemporary Anglo-American philosophy of mind it has once again become commonplace to assume that persons are no more than a special case in the philosophical analysis of consciousness, rationality, and psychology generally.

In The View From Nowhere, for example, Thomas Nagel approaches the topic of personhood by starting with the relation of consciousness to an ‘objective’ world which ideally would include minds among other things. Although this analysis leads to a more objective understanding of mental life itself and "consciousness in general," regarded as a

---

1138 This is precisely what the impersonality of Aristotelian moral psychology leaves out, on Frankfurt’s account. See Frankfurt, "On the Usefulness of Final Ends," p.5 (which I discuss in Chapter II).

1139 Thomas Nagel, The View From Nowhere (Oxford University Press, 1986).
realm of its own.\textsuperscript{1140} Nagel soon encounters, within the realm of consciousness, a more specific set of problems revolving around what we might call the first-person sense of self. Nagel introduces the cluster of questions involving the publically identifiable person and the private "I" through what he characterizes as the "particularly difficult mental concept..of personal identity over time."\textsuperscript{1141} The subsequent analyses he gives of the transcendental self, agency, and autonomy are all approached as special cases within the broader problem of conscious mental subjectivity generally (including that of animals). As a result, in his view an account of the meaning of first-person subjectivity is important mainly because of the difficulties it poses for the governing problem of reconciling the mental and physical sides of reality.

On the contrary, we have seen that Frankfurt’s focus on volitional identification points towards much more radical implications. Personhood involves a projective power of self-motivation, and a corresponding capacity for reflexive volitional motivation whose irreducibly intrasubjective character is the ground of an ineliminable sense of individuality within the self that goes beyond mere self-consciousness. Our capacity for identification gives us an active sense of owning and being responsible for our own motivational character. This kind of volitional individuality is not adequately captured in terms of conditions expressing our abilities in cognition and sentience. Hence volitional identification is crucial evidence in favor of the existentialist’s insistence on an essential individuality in personhood that transcends the natural kind notion of a rational animal.

Suppose we even granted, counterfactually, that Dennett’s various types of purely pragmatic "intentional stance" sufficiently captured the consciousness and rationality

\textsuperscript{1140}Ibid, pp.17-27.
\textsuperscript{1141}Nagel, p.32.
characteristics of persons, or that we assume for the sake of argument that we only care about *attributing* these features to agents from the third-party perspective. Even then, Dennett’s reconstruction of volitional identification would fail, because it fails to capture the immediately first-personal significance of the phenomena. And the most immediate implication of this is that personhood *cannot* be adequately evaluated from the pragmatic, third-personal perspective of "heterophenomenology." For what is ‘peculiar’ to personhood is a capacity for authentic identification which is *essentially* first-personal and irreducible to any set of conditions whose ‘objectivity’ depends on third-person neutrality with respect to individual identity. Identification and alienation are quasi-indexical phenomena, to use Hector-Nera Castenada’s term: their meaning is essentially dependent on the intrapersonal context of the individual who *enacts* them, even though they may have an analogous sense for all possible persons (as I suggested in the Introduction §1).

Pascal is not wrong when he argues that a man may lack hands, feet, or even a head, but says "I cannot imagine a man without thought; he would be a stone or an animal." As we have seen, the ability to think rationally and to bring rational reflection to bear on the formation of one’s will is an indispensible condition for moral responsibility and personhood in Locke’s ‘forensic’ sense. But Pascal is wrong to conclude with Descartes that mind and its modes of thought are the whole essence of personhood. A special kind of self-consciousness involving immediate reflexive self-awareness, and a capacity for rationality (including strategic calculation), as well as reflective interpretation of one’s motives may all be necessary conditions of personhood. But they are not sufficient conditions, and more importantly, no additional conditions of the same sort can capture the volitional aspects of personhood they leave out.

---

Most radically, then, the phenomenological project that Frankfurt started shows that the will which is essential to persons—or selves capable of moral responsibility—is something irreducible to other states of consciousness or rational cognition. As Heidegger holds, then, what persons are cannot be adequately explained in terms of classical metaphysics of kind-essences at all: persons who can identify themselves with different projects and other first-order states of their own motivational character show that the metaphysics which first led to the incorporation of volition into nous must be abandoned.

Our phenomenology of identification points towards these radical conclusions, and thus ultimately towards an existentialist conception of persons. Such a conception, however, requires one other feature which neither Frankfurt’s nor my own account of identification in terms of projective motivation has yet reached: namely freedom in the libertarian sense. The account of the will developed in this dissertation will provide an adequate basis for a existentialist analysis of personhood only if it can be shown the projective motivation and volitional identification are free in a suitably libertarian sense, and that liberty of identification can be convincingly explained in a way that avoids familiar problems of arbitrariness in decision. These tasks are left for future work.
[The official dissertation ended with Chapter Five, but I had originally prepared a substantial sixth chapter on freedom, which I include here in its final draft form.]

CHAPTER VI

Freedom of the Higher-Order Will
Contents of Chapter VI

VI.1 Introduction: The Libertarian Approach to the Origin of Identification

VI.2 Can We Leave Out Identification?: Fischer’s Account of Action-Responsibility in terms of Reasons-Responsiveness

VI.3 Frankfurt’s Theory of ‘Free’ or Morally Imputable Action refers implicitly to the notion of volitional identification

VI.4 Frankfurt’s Theory of ‘Free Will’ is compatible with liberty of identification

VI.5 Moral Responsibility for Character Requires Liberty of Identification
   VI.5.1 Internal Compatibilism cannot avoid the Action / Character Distinction
   VI.5.2 The Problem with Frankfurt’s ‘Quick Argument’ that character-responsibility does not require liberty of identification
   VI.5.3 The initial presumption in favor of the Requirement of Liberty of Identification (RLI) as a principle for character-responsibility.
   VI.5.4. The Impossibility of True Frankfurt-Counterexamples to RLI

VI.6 An Existentialist Response to the Challenge of Volitional Necessities
   VI.6.1 Volitional Inertia vs Volitional Necessity in Frankfurt’s Examples
   VI.6.2 A Sketch of Constrained Liberty in the Highest-Order Will
   VI.6.3. The Phenomenological Problem of Individual Volitional Essences
   VI.6.4 Between Frankfurt and Sartre: Neither Volitional Essence nor Absolute Groundless Freedom

VI.7 Liberty without Arbitrariness: An Existentialist Response to Four Objections
   VI.7.1. Why Existentialists Must Reject ‘Linkage’ between the two aspects of choice
   VI.7.2. Four Objections Against ‘Autonomy’ As Liberty
   VI.7.3. Outline of an Existentialist Defense Against these Objections
   VI.7.4. Conclusion: The Middle-Soul Concept of Will
VI.1 Introduction: The Libertarian Approach to the Origin of Identification

In Chapter IV, I argued that an account of identification in terms of second-order projection could explain the required qualitative difference between the states constituting volitional identification and other states to which identification may or may not be conveyed. I also argued that this approach could be extended into a broadly existentialist conception of decisive identification or the highest-authority will, although I did not work out that conception in much detail. Chapter V defended this approach to identification and identification, as I labelled them respectively, against at least some of the main alternative accounts proposed by others in the last twenty years.

However, the conclusions of Chapter IV left undecided whether or not, to capture the qualitative distinctiveness of volitional identifications, we also need to recognize conditions on their intentional history or causal origins. That is, it left undecided how best to answer the different aspects of Objection B1 to Frankfurt’s initial model. The arguments in Chapter V have addressed part of this problem: they suggest that accounts which rest the qualitative difference of identification-conveying states solely on their being the product of processes that are reasons-responsive in various senses, as opposed to deriving from direct manipulation or deficient deliberation or insufficiently comprehensive reflective evaluation, are overly cognitivist. In this crucial respect, such rational-origin models of identification are inferior to the projective model, which emphasizes the origin of identification in the agent’s distinctively volitional (though not arbitrary) projection of ends concerning her own outward character (i.e. concerning her first-order motives and dispositions to act). The projective approach may thus yield a more satisfactory resolution to some of the issues involved in Obj.B1.ii.

But this still leaves unresolved the question of whether, if identification is the volitional
commitment or devotion to attaining a particular outward character, such commitments
themselves must be freely made in the libertarian sense to count as volitional or as projected
by the agent for himself. This chapter will argue that this resolution to the problem of
Obj.B1.i enjoys \textit{prima facie} attractiveness, and that it can be defended against a host of
possible objections, ranging from counterfactual intervener-type counterexamples to the
claim that libertarian freedom at this level entails an arbitrariness destructive of selfhood.

The possibility of a libertarian approach to identification was already introduced in
Chapter IV. As we saw in IV.4.1, Frankfurt cannot simply presuppose that any qualitative
account of identification that will answer the \textit{regress} and \textit{ab initio} problems (Obj.B2-B3)
will therefore render the causal and intentional history of identification-conveying states
irrelevant. Thus Frankfurt cannot dismiss the problem at stake in Obj.B1.i. The possibility
of arguing that only a libertarian historical conception of identification-conveying states fully
accounts for their qualitative difference thus remains open in principle.

Yet it is important at the outset to distinguish two different senses in which one could
claim that libertarian freedom should be considered an integral part of initiating volitional
identifications, understood now as self-referential or reflexive projections. This thesis could
be meant either in an \textit{absolute} metaphysical sense, e.g. as the thesis that agent-causation
must exist for identification to be possible, or as a \textit{practical presupposition} that the agent is
at liberty to form different identifications. On the latter alternative, the agent must regard
herself as having such ‘liberty of the higher-order will,’ but this does not rule out the
possibility that determinism might still be true for all we know (see the discussion of
Kantian libertarianism in Chap. II §5.6). As we will see, Frankfurt is concerned to deny that
identification requires liberty in \textit{either} of these senses. But my argument in this chapter will
focus only on defending the second or weaker version of the thesis that identification
involves liberty. For it is the idea that we must regard ourselves as being able to identify otherwise that enjoys the *prima facie* plausibility from which my argument against Frankfurt (and others) on this issue begins.

VI.2 Can We Leave Out Identification?: Fischer’s Account of Action-Responsibility in Terms of Reasons-Responsiveness

A complete treatment of the multiple debates surrounding free will and compatibilism today would begin by considering the current state of play in several fields, including: (1) problems with ‘direct’ arguments for the incompatibility of determinism and moral responsibility; (2) prospects for Van Inwagen-style transfer-principle arguments for the incompatibility of natural determinism and the freedom to do otherwise; and (3) the very complex debate over whether Frankfurt-style counterfactual intervener cases show that moral responsibility for actions, omissions, and consequences of acts does not require the freedom to bring about alternative possible actions, decisions to act, or consequences of action. I hope to address these three important topics in future work, but in this chapter, I will instead focus directly on the issue of liberty in the higher-order will.

In my view, the jury is still out on whether or not Frankfurt-style counterexamples disprove the Principle of Alternate Possibilities as Frankfurt formulated it for first-order actions, and similar principles for omissions and consequences of our acts. If Frankfurt-style counterexample *do* refute these PAP-like principles, this does not immediately prove that responsibility is compatible with causal determinism, but it supports the conclusion that moral responsibility for first-order acts, omissions, and consequences must be constituted by

---

1143Frankfurt, "Alternate possibilities and moral responsibility," reprinted in *The Importance of What We Care About*, p.1: "This principle states that a person is morally responsible for what he had done only if he could have done otherwise. Its exact meaning is subject to controversy, particularly concerning whether someone who accepts it is thereby committed to believing that moral responsibility and determinism are incompatible."
some set of *actual sequence* conditions. By itself, however, this conclusion would still tell us little about what kind of actual sequence model will be most phenomenologically adequate. In this section and the next, I will compare and contrast John M. Fischer’s actual sequence model with Frankfurt’s, and argue that the latter has stronger phenomenological appeal.

In his book, *The Metaphysics of Free Will* and earlier articles, John Fischer presents a model for responsibility in terms of "guidance control" which, unlike "regulative control," does not require libertarian freedom. But unlike Frankfurt’s, Fischer’s actual-sequence model tries to explain guidance control *without* referring to the phenomena of volitional identification. This implies a different way of dealing with well-known problems for early compatibilist models, such as persons who do what they choose, but whose choices result from coercion, compulsive disorders, psychotic fears, brainwashing, or other forms of responsibility-undermining obsession, manipulation, and conditioning. While Frankfurt attempted to resolve such problem by referring to the role of higher-order volitions, Fischer tries to extend compatibilist explanations to these problem cases by developing the notion of a reasons-responsive mechanism that operates to produce our actions in the actual sequence of events. As I understand it, such a mechanism is ideally supposed to capture what it is to make choices *sanely* or voluntarily. For example, when a person acts on "a literally irresistible urge, he is undergoing a kind of physical process that is not reasons-responsive," and so is not responsible for his act.¹¹⁴⁴

Specifically, Fischer proposes four conditions for action-responsibility. (1) First, the agent’s action must result from a mechanism that is *weakly reasons-responsive* (WRR): "under weak-reasons-responsiveness, there must exist some possible world in which there is

sufficient reason to do otherwise, the agent’s actual mechanism operates, and the agent does otherwise."\textsuperscript{1145} In this case, it is the fact that there is such a world, rather than its similarity to the actual world, that counts. The opposite is true for strong-reasons responsiveness (SRR), which "obtains when a certain kind K of mechanism actually issues in an action and if there were sufficient reason to do otherwise and K were to operate, the agent would recognize the sufficient reason to do otherwise and thus choose to do and do otherwise."\textsuperscript{1146} SRR is satisfied only if in \textit{any} world —including those closest to the actual world— in which SRR operates, a similar choice-circumstance occurs, and yet the reasons for action dictate a different action, the agent acts as reason dictates. As Fischer says, this seems to be too strong a condition for moral responsibility, however, since it would inappropriately exculpate me in situations where I knew I had sufficient reason to do otherwise but, through weakness of will, I choose the worse action.\textsuperscript{1147} Under WRR, I can act on an option I know to be worse and be responsible for it, as long as "there are some circumstances in which I would respond appropriately to sufficient reasons."\textsuperscript{1148} For example, an agent might steal a book knowing this is wrong, and be responsible for it: but if he would persist in stealing a book "even if he knew that by so acting he would cause himself and his family to be killed, then the actual mechanism would seem to be inconsistent with holding him morally responsible for his action."\textsuperscript{1149} 

Fischer’s thought here is clearly that in cases where some unusual factor like obsession or hypnosis etc. is influencing the agent, he would do what he does in \textit{every} possible scenario.

\textsuperscript{1145}Ibid, p.166.
\textsuperscript{1146}Ibid, p.164.
\textsuperscript{1147}Ibid, p.166.
\textsuperscript{1148}Ibid, p.167.
\textsuperscript{1149}Ibid, p.167.
presenting the same choice-circumstances. This might be a bit strong, however: suppose a psychotic actor would refrain from stealing the book if he knew that he would be caught and re-institutionalized, and that this happens to be sufficient prudential reason not to steal it, but that in no other circumstance (no matter how extreme) would he refrain from stealing it, since returning to the mental hospital is the only thing he fears strongly enough to overcome his obsession with the book. This single instance is certainly not enough to make the mechanism on which he acts reasons-responsive in the relevant sense, although it is enough to make him satisfy the strict definition of WRR. Surely then, to be sufficient reasons-responsive, he must act otherwise in enough of the cases where the reasons to do otherwise are strong enough.\footnote{Fischer recognizes this problem (ibid, p.243, note 8) and addresses it in his new book with Mark Ravizza. Van Inwagen makes a bit too much of this problem in his review of The Metaphysics of Free Will (see note 17 below). For as I suggest here, what is needed is a more nuanced conception of sanity, not alternative possibilities.} But how exactly to draw these boundaries to define sane action will be difficult to determine in any terms universally applicable to very different types of choices: a prudential judgment is implicated here. As Fischer says, on his approach, "Moral responsibility requires some connection between reason and action, but the fit can be quite loose."\footnote{Ibid, p.168.} But I think it cannot be quite as loose as the formal definition of WRR implies, and exactly how loose it needs to be instead may be impossible to specify in a short set of formal conditions. This does not show that Fischer’s reasons-responsiveness condition is unworkable, though it indicates how complex the notion of sane action may actually be.

(2) Second, it must be the relevant mechanism underlying the action that is weakly (or I suggest: sufficiently) reasons-responsive. To pick out the relevant mechanism, we have to ignore ones that are defined by relation to a specific action-outcome (analogously to ‘soft facts’): the relevant mechanism must be "temporally intrinsic," meaning that "if a
mechanism M issues in act X, then M is relevant to the agent’s guidance control of X only if M’s operating does not entail that X occurs.”\textsuperscript{1152} In addition, the relevant mechanism must be the one in terms of which it is most intuitive to explain the action: Fischer assumes that “for each act, there is an intuitively natural mechanism that is appropriately selected as the mechanism that issues in an action.”\textsuperscript{1153} The idea here is like that of intentional explanation: it must be possible to give some coherent narrative in terms of reasons, motives, considerations, etc. that make intelligible why A did X. This could be fleshed out in terms of what the agent’s own honest and reflective account of her act would be, or (to deal with the problem of unconscious factors) in terms of what an ideally informed observer or omniscient introspector of minds would say about the action. However, Fischer suggests that he can if necessary avoid the problem of filling out this condition by saying that "an agent exhibits guidance control of an action insofar as there is no actually-operative (temporally intrinsic) mechanism issuing in the action that is not reasons-responsive. This alternative obviates the need to select one mechanism as the relevant one."\textsuperscript{1154} This is unconvincing, however, since it would lay Fischer’s account open to actual-sequence overdetermination objections. Imagine a case where multiple sequences operate in an agent to cause him to read a certain novel: he heard it was entertaining, he wants to improve his reading, he has reason to think the book will educate him about important historical topics, etc., and (by chance) he is irrationally attracted to the picture on the book’s cover. Surely it is still (mildly) laudable for him to read the book, although one non-reasons-responsive factor contributed to his reading it. Again, then, a prudential judgment about what mechanism(s) were most important in the causal or intentional account of an agent’s action.

\textsuperscript{1152} Ibid, p.172.
\textsuperscript{1153} Ibid, p.173.
\textsuperscript{1154} Ibid, p.173.
seems unavoidable on the approach Fischer advocates.

(3) Third, although WRR requires that "there must be a possible scenario in which the same kind of mechanism operates and the agent does otherwise," the *sameness of kind* of mechanism "need not require sameness of all details, even down to the micro-level." Thus if causal determinism holds, the relevant mechanism is not defined in terms of the particular causal process that nomologically necessitates the action. This is important, of course, because otherwise causal determinism would imply that the kind of alternate scenarios needed to establish weak (or sufficient) reasons-responsiveness never exist. On Fischer’s account, there only need to be scenarios in which the same *sort* of mechanism—not the same mechanism-particular—operates and the agent follows different reasons and does otherwise. As Fischer says, this yields *semi-compatibilism*: "moral responsibility is compatible with causal determinism, even if causal determinism is incompatible with freedom to do otherwise" (i.e. libertarian freedom or regulative control). Note that point about "sameness" of mechanism again suggests, however, that "mechanism" is really more of an intentional process, rather than the causal process in which a set of intentional functions are realized. Moreover, judging when causally different mechanisms are sufficiently similar in the *intentional* sense to count as "the same mechanism" will also be a matter of prudential judgment. So whether the agent’s deliberative or motivational process was sensitive to reasons—meaning that the *same* process would have led to a different conclusion in some of the circumstances (suitably selected) where there are different considerations—will depend in part on what different types of deliberative or motivational processes our theory of reason and motivation recognizes (and as functionalism implies, this

---

1155 Ibid, p.179.
is not just a causal matter).

(4) Finally, Fischer adds the caveat that a person can be responsible for acting on a non-
reasons-responsive mechanism if that mechanism itself issued from a weakly reasons-
responsive process: "we must be able to trace back along the history of the action to a point
(suitably related to the action) where there was indeed a weakly-reasons responsive
mechanism." Here again, as the parenthetical remark indicates, whether we can trace
back to a responsible choice will depend on both practical judgment of the case at hand, and
on what different types of intentional accounts or motivational narratives our theory
recognizes.

The extent to which this approach seems promising, then, is probably proportional to its
similarity to the kind of nuanced account of voluntary action that Aquinas and others in the
same tradition developed. But this first full sketch of the reasons-responsiveness account
has been subject to numerous proposed counterexamples. In a review of Fischer’s book,
Peter Van Inwagen argued that there are clearly cases in which the fact that a result would
obtain no matter what exonerates the agent from responsibility when this result follows from
his action. Fischer has arguably done a good job of responding to such cases by
extending his model to explain that one is responsible for consequences only when one’s
bodily movement issues from an appropriately reasons-responsive mechanism, and "the
process going from the pertinent bodily movement to the event in the external world is
suitably sensitive to the bodily movement" (which is not the case in Van Inwagen’s ‘plague-

\[\text{\footnotesize 1157} \text{Ibid, p. 176.}\]
\[\text{\footnotesize 1158} \text{To this extent, I agree with Eleonore Stump’s frequent comparisons of Fischer’s and Aquinas’s treatments.}\]
\[\text{\footnotesize 1159} \text{Van Inwagen, "Fischer on Moral Responsibility," The Philosophical Quarterly, Vol.47 (1997): 373-81. For example, Van Inwagen describes a case in which an officer is supposed to take a serum upriver to a plague-stricken village: he gets drunk and misses the boat, but the boat sinks right after leaving the dock. The officer is not responsible for the fact that hundreds die who would have been saved had they received the serum.}\]
stricken village’ case and similar examples.\textsuperscript{1160}

But Fischer has more difficulty with another sort of case. Stump has argued that in cases where an agent A acts on a mechanism that is controlled by others who are themselves acting on reasons-responsive mechanisms, A will satisfy Fischer’s conditions and yet obviously fail to be responsible.\textsuperscript{1161} Recently, using an example from Robert Heinlein’s \textit{The Puppetmasters}, Stump argued that an agent could act on a mechanism controlled so as to be reasons-responsive by an alien consciousness, and obviously not be responsible for action that is so directly manipulated by the alien:

What Fischer needs, then, is some way of specifying that the weakly reasons-responsive mechanism on which an agent operates \textit{belongs to that agent}. It isn’t enough for moral responsibility that the agent does what he does under the guidance of a weakly reasons-responsive mechanism. The mechanism has to be the agent’s own....in some suitably strong sense.\textsuperscript{1162}

The problem is that without identification, there is no way to link the mechanism to the agent. As R. Jay Wallace also argues, the problem lies in Fischer’s assumption that "moral accountability" can be explained through the modal properties of the action-causing mechanisms:

This approach brings an objectifying, third-personal vocabulary to bear on phenomena that have their natural place within the deliberative perspective of practical reason, with the result that the intuitive locus of responsibility, the person, seems to drop out of view.\textsuperscript{1163}

In their new book,\textsuperscript{1164} Fischer and Ravizza introduce a model of moral responsibility based on the


notion of a "moderately reasons-responsive mechanism" that must produce the outcomes it does because of its recognition of reasons (rather than just matching what someone acting on reasons would do). This eliminates many of the counterexamples to Fischer’s earlier model. But as Stump points out, the man in her example from Heinlein meets the new conditions as well: "when Sam under the control of the alien does A, it is because there is a reason for Sam to do A, and the mechanism controlling Sam is responsive to that reason for reasons of its own." To eliminate this kind of counterexample, Fischer and Ravizza have to stipulate that the mechanism on which the agent acts count as her own, and that she takes responsibility for this mechanism by seeing herself as the source of actions it produces and accepting this as a basis for reactive attitudes directed at her.

But this is inadequate, as another example borrowed from Frankfurt will show. As Fischer’s analysis implied, when someone strongly addicted to cocaine takes some of the drug, this is a paradigm case of acting on a compulsive rather than reasons-responsive mechanism. Yet suppose that, through another series of internal intentional states, the addict wills to act on such addictive motives, or identifies with being an addict. Which is the relevant mechanism? As this example suggests, whether the agent admits that the internal process or sequence that is the source of her acts is ‘her’ in the sense necessary to make her see herself as responsible for its products will depend precisely on whether she identifies with it. The willing addict sees the compulsive mechanism as her own (and does so in a way that is not simply part of the compulsion). That Fischer has overlooked this particular difficulty for his approach is suggested by one of his earlier essays, where he says that the "happy addict" may be responsible for taking the drug since "his addiction plays no role in his

---

1165 And this modification addresses one of R. Jay Wallace’s criticisms in his review of The Metaphysics of Free Will, p.159.


1167 Ibid, p.192. Stump argues that Sam in her example could have all these attitudes and yet be deceived about being responsible for his actions. This is not so troubling for Fischer and Ravizza, however, since they admit that we can be fallible in taking responsibility for ourselves, just as we can be fallible in judging whether others are responsible.
decision and action" of taking the drug.\footnote{1168 Fischer, ed., *Moral Responsibility*, Introduction, p.42.} There is no mention of identification here, and this suggests that Fischer would account for the willing addict as one who is acting on a reasons-responsive mechanism, although a non-reasons responsive mechanism also determines her to the same action. But the addictive desire is not a counterfactual intervener: it is a simultaneous contributor to the action. Thus to accommodate this example in terms of his theory, Fischer needs to say that the relevant mechanism in the willing addict is the non-addictive one. But on what grounds can we say this? If the criterion for relevance is causal, then the addictive mechanism may be the relevant one: for as Frankfurt emphasizes, the addict’s identification with her desire may not be what causally brings about her action at all (the addictive urge may be the only motive playing a causal role). But if the criterion for relevance concerns the qualitative difference between types of intentional states involved in the mechanism, then ‘relevance’ is simply equivalent to the type of agent-authority we mean by ‘volitional identification.’

By contrast, consider someone who unwillingly acts out of a desire to help her sister. Suppose that this desire is not unconditional: it is a reasons-responsive mechanism. Further, assume that this mechanism’s role in her action is not determined by aliens or external forces. On Fischer’s account, she should be praiseworthy for helping her sister in a time of need, but yet she does not identify with acting on the considerations and interests that moved her in the mechanism we’ll call ‘concern for sister’s well-being.’ She recognizes that in a sense, the action issues from within ‘her,’ but she gives this action and the mechanism that produced it no place in her heart, in her deeper self, and thus will accept at most a diminished responsibility for it. Here again, the relevant mechanism for determining the agent’s responsibility may not be the one that plays the decisive causal role in her action.

For reasons of this sort, I am pessimistic about the possibility of constructing any actual-sequence
model of moral responsibility that does not make at least implicit reference to the phenomena of identification. Yet unlike the views examined in Chapter V, Fischer rejects identification as a crucial component in the account of moral responsibility because he takes it to be what he calls a "mesh" condition: "A mesh theory isolates some harmony within the internal psychological economy of the agent and posits that this is freedom-relevant condition for moral responsibility." Fischer says that such mesh-conditions are too a-historical to be of any help:

...the obvious problem with such an approach is that the selected mesh [or harmony] can be produced by intuitively responsibility-undermining processes. And the same general sort of problem afflicts even the refined versions of Frankfurt’s theory....

...The general problem with such theories is that they are purely structural (focusing on internal relationships rather than the relationship between the agent and the external world) and a-historical (focusing on the current time-slice rather than the process which issues in the action).\textsuperscript{1169}

The same objection applies to Watson’s "multiple source" version of mesh theory that equates identification with acting on preferences approved by one’s values: the mesh between "elements of different preference systems may be induced by electronic stimulation, hypnosis, brainwashing, and so on."\textsuperscript{1170} Fischer’s historical approach to responsibility in terms of reasons-responsiveness in the process leading up to action is inspired precisely by this deficiency with Frankfurt’s identification-based approach: "The moral I draw from these considerations is that an adequate theory of moral responsibility will attend to the history of an action, and not simply to its current time-slice characteristics. Further, it will require more than simply some internal arrangement of elements of an agent’s mental economy."\textsuperscript{1171}

---

\textsuperscript{1169} Fischer, \textit{The Metaphysics of Free Will}, p.208.

\textsuperscript{1170} Ibid, p.209.

\textsuperscript{1171} Ibid. Fischer points out that the unlike purely structural models, his historical or actual-sequence approach is the \textit{kind} of model that can deal with such problems, though his own version of it as formulated may still be subject to the objection that a weakly reasons-responsive mechanism could be artificially induced (p.209). In his review of Fischer’s book, Warfield takes Fischer to task for this, but fails to note the superiority of the historical approach over the structural approach relative to this kind of problem. The debate between (continued...)
In my view, there is clearly something compelling in Fischer’s objection to Frankfurt here, but to see what it is, we have to view Fischer’s point in terms of the foregoing analysis in Chapter IV. From that perspective, Fischer runs together the problems I distinguished as Obj.B1 and Obj.B4, and he fails to note that if identification is understood projectively (or in terms of Frankfurt’s care theory) it has a historical dimension: it is a *dynamic* condition rather than a time-slice one. But still, this new account does not focus on the causal history of the persisting states that constitute care, and so it remains *structural* in Fischer’s sense. And it does not place any limits on the prior sequence of intentional states that might lead up to the strong evaluations and decisions constitutive of projective care. So the refined account still leaves open the troubling possibility that the persisting formation of states and dispositions concerning one’s motives that constitutes volitional identification could result either from direct causal manipulation, or from obsession, coercive threats, conditioning etc. As Paul Stern similarly complains,

> Frankfurt’s view construes freedom of the will exclusively in terms of the effectiveness of one’s second-order volitions in determining first-order desires, and it thereby neglects the problem of whether the agent’s adoption of these second-order volitions is itself free or the product of ‘forces’ alien to him. 1172

Fischer and Stern are right to stress this, and as I will argue below, only recognizing that the source of identification involves libertarian freedom can adequately resolve this problem.

But Fischer’s mistake lies in treating the "source-problem" for non-libertarian theories of identification (however refined) a reason to eliminate the *phenomenon* of identification from its crucial role in our experiences of moral responsibility. In other words, what his objection shows is that we need an adequately historical account of identification, not that we can do without the

---

1171(…continued) Fischer and libertarianism comes down to whether only a *libertarian choice* in the process leading up to an action could rule out such manipulation counterexamples. Fischer’s type of compatibilism has a chance against the libertarians in this debate, whereas a purely structural account such as Frankfurt’s does not.

1172Stern, Translator’s Introduction to Tugendhat’s *Self-Consciousness and Self-Determination*, p.xxi.

1173As Fischer calls it: see *The Metaphysics of Free Will*, p.250, note 17.
concept of identification in explaining responsibility for first-order action. Thus although purely structural accounts of identification may fail, still, no historical account of moral responsibility (whether it requires libertarian freedom or merely "guidance control") may be convincing unless it includes states of identification in this historical processes behind an action. In this respect, as the next section will show, Frankfurt’s approach to moral responsibility is intuitively more appealing, although he continues to give unduly ahistorical accounts of identification and he remains surprisingly unwilling to enter into the debate about what kinds of limits must be placed on the source of states for them to constitute identifications.

What this suggests is that the right approach will combine Fischer’s emphasis on historical sources with Frankfurt’s emphases on identification and the idea that the factors accounting for responsibility need not be the ones causally relevant for the agent’s outward actions. This is the hybrid approach I attempt to sketch in this chapter.

VI.3 Frankfurt’s Theory of ‘Free’ or Morally Imputable Action refers implicitly to the notion of volitional identification

My aim in this section is to show that Frankfurt’s own account of responsibility for (first-order) actions is different from Fischer’s in a crucial respect. Frankfurt has consistently maintained that "freedom of action" —which is defined as the kind of freedom that makes actions imputable— has two basic conditions. As we will see, the second condition involves implicit reference to the phenomenon of volitional identification, and it is this feature which distinguishes Frankfurt’s analysis of responsibility from other actual-sequence models.

(1) The first condition is the one stressed by a long line of compatibilists stretching from Hobbes
I owe thanks to Michael McKenna for some tips on the background of this condition, which he has treated in his dissertation.


Frankfurt, "What we are morally responsible for," p.100. For example, an unwilling addict does not 'act freely' when he is about to move his hand to shoot up with heroin, but is stopped by a sudden seizure or by someone else's intervention (though at a deeper level he may be glad that this happened).

This is confirmed by Frankfurt’s analysis in his earlier paper on "The Problem of Action," American Philosophical Quarterly, 15 (1978), reprinted in The Importance of What We Care About: 69-79. In this paper, Frankfurt argues that "action is intentional movement" (p.73) or purposive behavior (p.74), which he distinguishes from intentional action or self-conscious identification with the intention that will guide movement. Thus ‘intentional action’ is analogous in this essay to second-order volition in the "Freedom of the Will" paper. In this terminology, the unknowing addict acts merely through behaving according to his first-order will: "His movements are not mere happenings, when he takes the drug because he cannot help himself. He is then performing the very same action that he would have performed had he taken the drug...with the illusion that he might have done otherwise" (p.76). Thus one does what he intends, in Frankfurt’s sense just when one acts: "The assertion that someone has performed an action entails that his movements occurred under his guidance, but not that he was able to keep himself from guiding his movements as he did" (p.77).

As Frankfurt makes clearer in his response to Van Inwagen, for him Condition (1) means that a person’s behavior is guided by his (first-order) intention: the 'action' imputable to an agent consists of his "bodily movements" made 'with certain intentions or expectations." In other words, one does what one wants when one’s behavior is guided by one’s first-order will.

As Frankfurt suggests, however, what this guidance amounts to is open to some question. On the one hand, "guidance" could be construed as a modal concept: a bodily movement is "guided" in the actual sequence only if in counterfactual sequences where the first-order will is different, appropriately different bodily actions would also follow. Condition (1) would then be similar to the hypothetical criterion of freedom that Roderick Chisholm derives from G.E. Moore: "If he had..."
chosen to do otherwise, then he would have done otherwise” (which is of course still compatible with lacking even the nomological possibility of choosing otherwise). On the other hand, if "guidance" only requires an accidental accord between one’s intentions and one’s bodily movements, then Frankfurt’s condition (1) is a Schlickean principle implying only that no impediments prevent one from making bodily movements in according with one’s intentions in the actual sequence.

This second interpretation seems likely, because in typical ‘Frankfurt-style’ overdetermination counterexamples to PAP, where Jones wants ‘on his own’ to do X in the actual sequence, which happens to be what the manipulating devil or neurologist (D/N) also wants him to do, and in the counterfactual sequence where Jones wants to do Y instead, he would not have done Y because the D/N would make him do X instead. Thus Jones lacks Moorean hypothetical freedom of the will but is still responsible for his act in the actual sequence, according to Frankfurt, and so his action must satisfy condition (1).

However, there are reasons elsewhere in Frankfurt’s work to favor a more Moorean reading of condition (1). In his paper on "The Problem of Action," Frankfurt argues that causal theories of action are subject to counterexamples. Instead, he suggests that an action consists in bodily

---


1180 Assuming that if there are no accessible possible worlds in which the counterfactual condition is satisfied, then the whole counterfactual is vacuously true.

1181 See Moritz Schlick, General Theory of Knowledge, tr. Albert Blumberg (Open Court, 1985): "We call human behavior free when it proceeds from motives in a normal way, without being inhibited by obstacles that lie outside the nature of the person in question" (p.380). And given Schlick’s psychologistic reduction of the meaning of modal concepts, "proceeds" here cannot be understood as "caused" in any strong nomological sense.

1182 Frankfurt’s argument on this point has recently been defended by Stewart Goetz in his paper, "Libertarian Choice," Faith and Philosophy 14.2 (April 1997): 195-211, pp.197-199.
movements that are "under a person's guidance" as they occur.\textsuperscript{1183} He insists that "the question of whether or not movements occur under a person's guidance is not a matter of their [causal] antecedents," but of whether they "cohere in a pattern which strikes us as meaningful" as they unfold.\textsuperscript{1184} But if this is in part a matter of whether a person adjusts her movement-in-process to compensate for contingencies as they arise, thus keeping the movement 'on course' to its intended result, then guidance must involve counterfactual conditions after all.\textsuperscript{1185} Note that there is also a way of interpreting the Frankfurt's own counterexamples to PAP that fits with this more Moorean conception of guidance: the D/N in the counterfactual sequence not only makes Jones perform X-movements, but also makes Jones guide them by an intention to X (so that Jones does X in Frankfurt's sense of an action). Note that this is compatible with the assumption that Jones has hypothetical freedom of the will: if Jones intended to do Y instead, he would have guided his movements accordingly. This result follows from the stipulation that the intervener only intervenes to control intentions: thus if Jones intended Y, the intervener did not intervene.

If instead we say that in the counterfactual sequence the D/N would only make Jones perform X-movements, while Jones still intends to do Y, then in the counterfactual sequence Jones does not "act" at all in Frankfurt's sense, since it is the D/N's intentions and not Jones's that guide Jones's bodily movements. Still, in that case, Jones satisfies a slightly modified 'hypothetical' or

\textsuperscript{1183}Frankfurt, "The Problem of Action," p.72.
\textsuperscript{1184}Ibid.
\textsuperscript{1185}This would be regard "guidance" as a contingency-responsive mechanism in which the agent would complete the same action according to her intention in at least the closest alternative possible worlds in which unexpected events occur during the course of the action. Note that this corresponds with the third of the three aspects of reasons-responsiveness or guidance control in Fischer's account: "the connection between choice and action" (\textit{The Metaphysics of Free Will}, p.165). But Frankfurt requires strong responsiveness to contingencies for this aspect. There is good reason to think that we must move in this direction because of the problem of "tropistic behavior," which seems to cohere in an intention-guided pattern, but is revealed as entirely unguided when it fails to compensate for interruptions in the routine. The mechanisms behind such behavior are insufficiently responsive to unexpected contingencies to count as guidance mechanisms. See Daniel C. Dennett, "Mechanism and Responsibility," reprinted in \textit{Free Will}, ed. Watson, 150-173, esp. p.162, where Dennett discusses the tropistic behavior of the \textit{Sphex} wasp.
counterfactual Moorean condition: if Jones had intended otherwise than X, he would either have (a) acted according to his non-X intention, or (b) behaved in a way that is not an action at all. In this attenuated sense, Jones still has hypothetical freedom, and so condition (1) can still be construed as a type of Moorean condition for action-responsibility, without any negative impact on the counterfactual intervener arguments against PAP.

(2) A more precise interpretation of the details of the first condition for action-responsibility is not needed, however, to see how it works in conjunction with Frankfurt’s second condition, which is that the agent did not actively identify with another desire or intention than the one on which she acted. This is implicit in Frankfurt’s comment at the end of his "Alternate Possibilities" paper that to be excused from responsibility for his action, the person must have

...[done] what he did only because he was unable to do otherwise, or only because he had to do it. And we understand him to mean, more particularly, that when he did what he did it was not because he really wanted to do it.\textsuperscript{1186}

Here the notion of identification which Frankfurt developed two years later is already implicit. The point is that someone is not responsible for her act if she acts on a first-order desire that is alien to her.\textsuperscript{1187} For if she identifies with a desire contrary to the one on which she acts, then she was coerced by an impulse that was contrary to her will, and she acted as she did only because she was unable to do otherwise, for we intentionally guide our behavior according to a desire from which we inwardly dissociate ourselves only if we believe we have no other option. By contrast, if we identify with our


\textsuperscript{1187}Note that if we don’t interpret Frankfurt this way, we will end up in John Fischer’s puzzlement about the meaning of Frankfurt’s condition (2): “It is somewhat unclear what it would be for an agent to do something ‘only because he could not have done otherwise.’ Presumably, as I have suggested above, what distinguishes action from mere movements or mere events is that an action is preceded by an performed because of some appropriate pro-attitude (or perhaps volition)... It will follow that no action is performed solely because the agent could not have done otherwise; all actions, as opposed to mere events, will occur at least in part because of the pro-attitude” (\textit{The Metaphysics of Free Will}, pp.245-6, note 21. Fischer apparently does not see that in his 1969 paper, Frankfurt already has in mind that an action can be done intentionally, or with some pro-attitude, and yet not be identified with the agent (although Frankfurt does not formulate it this way until his 1971 paper).
action-maxim or intention, then we are definitely liable for what we do, whether or not we could have *acted* otherwise—as long our behavior follows this intention as required by condition (1). This reading accords explicitly with what Frankfurt says in "Three Concepts of Free Action:" "to the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions." As Fischer remarks, Frankfurt’s considered view is that "When one *identifies* (or perhaps ‘identifies decisively’) with the first-order desire on which one acts, this constitutes a sufficient condition for acting freely," and thus for responsibility for the action.

An example from the "Freedom of the Will" paper two years later also confirms this interpretation of condition (2): Frankfurt says that because the willing addict has made his addictive desire "his own" or identified with it, it is "not only because of his addiction that his desire for the drug is effective," and so he is morally responsible for taking it. Frankfurt also repeats this formulation from his "Alternate Possibilities" paper in his later essay on moral responsibility: "The fact that a person lacks alternatives does preclude his being morally responsible when *it alone* accounts for his behavior." But if his action follows from other "reasons of his own," then he is morally responsible for so acting, whether or not he could have acted otherwise. For Frankfurt, such "reasons of his own" obviously cannot consist of alienable desires, since otherwise the unwilling addict would be acting for reasons of his own, and thus be responsible (but he is not). What makes reasons *our own* is identification, which is always itself inalienable (as we saw in ch.IV).

Putting conditions (1) and (2) together, an agent ‘acts freely’ when he behaves (or guides his

---


1189 John Fischer, ed., *Moral Responsibility*, Introduction, pp.48-49. Note that the ambiguity in Fischer’s formulation reflects the confusion generated by conflating positive and decisive identification, as discussed in Chapter IV.

1190 Frankfurt, "Freedom of the Will...," p.25.

movements) according to his first-order will \(w_1 \cdot A\), and does not will \(w_2\) to act on some other motive contrary to \(A\) (since in that case he would not act on \(w_1 \cdot A\) for reasons of his own; he would act on his intention only because of the apparent impossibility of doing otherwise).\(^{1192}\) Note that this analysis allows us to hold a "wanton" responsible when her action follows or is effectively guided by her first-order will. As Frankfurt says, even non-human animals may often satisfy condition (1).\(^{1193}\)

Since condition (2) only requires the absence of a higher-order volition that would alienate the motive \(A\) acted on, it might also be thought that a wanton animal could satisfy condition (2) as well. However, for this reason, the idea that nothing more than satisfying (1) and (2) is needed for action-responsibility manifestly depends for its plausibility on the assumption that condition (2) only applies to beings that can form identifications (whether they do or not), and not to beings that are essentially wanton. Thus since Mary was capable of alienating the motive \(A\) on which she acted, but did not form any positive identification one way or the other, she may be regarded as tacitly identified with the motive \(A\) for the intention that effectively guided her action, and thus responsible for her act. My suggestion is that "neutrality" in the higher-order will could ground her satisfaction of condition (2) and thus her responsibility for her act in the way Frankfurt implies that it does only if

\(^{1192}\)This is similar to the result of Eleonore Stump's analysis in her essay, "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will," in Perspectives on Moral Responsibility, op. cit.: 211-234. Stump distinguishes three conditions for "acting freely;" "(1) having no external obstacles to doing what one wants to do, (2) having no internal obstacles to doing what one wants to do, and (3) having no external obstacles to willing what one wants to will" (p.215). Since Stump's (1) and (2) concern different kinds of obstacles that block the agent's first-order will from guiding her bodily movements, these conditions are merged in my 'Condition (1). And since Stump's (3) concerns obstacles that force an agent to have a different will \(w_1\) than she wants \(w_2\) to will \(w_1\), it is close to my 'Condition (2).' The difference is that my 'Condition (2)' does not require that the second-order volition be causally effective in determining the first-order will, but only that the first-order will not be in discord with an (unambiguous) second-order volition (if any). Since what freedom of will adds on Frankfurt's account is precisely this effectiveness of the higher-order will in guiding the intentions constitutive of bodily actions, Stump's analysis fails to capture the difference between 'acting freely' and 'willing freely' in Frankfurt's argument. This error carries over into her own revisions in Frankfurt's model, the second of which is a proposal to identify second-order volitions with effective second-order desires (p.217). This redeployment of Frankfurt's terminology has two costs: first, it eliminates Frankfurt's useful distinction between a desire merely to experience another desire and a desire to have a desire guide bodily action; and second, it obscures Frankfurt's insight that identification with a desire on which one would act does not require that this identification is effective in controlling the first-order intentions on which one acts.

\(^{1193}\)Frankfurt, "Freedom of the Will," p.20: "we recognize that an animal may be free to run in whatever direction it wants."
we can regard such neutrality itself as a kind of positively undertaken position. And neutrality can be a positive position only for beings who are not essentially wanton, but who can form identifications.\textsuperscript{1194}

Thus, although Frankfurt is not as clear on this point as he might be, his explanation of what it is to be responsible for one’s actions undoubtably depends on the capacity of persons for volitional identification with or alienation of motives that are ‘their own’ in the minimal sense. In this respect, the actual sequence model of action-responsibility that Frankfurt presents can be seen (when clarified and fully worked out) to differ significantly from Fischer’s proposed model. For Fischer, an intention is "ours" in the sense that makes us responsible for it because it has its causal origin in an actual sequence mechanism that is undistorted by the kind of external manipulation, delusion, compulsive disorder, phobia, or extreme irrationality (going beyond mere weakness of will) that would make this mechanism fail to be even weakly reasons-responsive.\textsuperscript{1195} For Frankfurt, on the other hand, our identifying with the motive on which we act (or minimally: refraining from alienating this motive) need have nothing to do with causing this motive to become our will, or bringing it about that the intention guiding our action springs from this motive. Thus our responsibility for an action A does not depend on the causal mechanism by which our A-intention originated in the actual sequence, but only with whether we (tacitly or actively) identify with this intention, or the motive that makes it practically intelligible. We act freely not because our act was causally controlled by a sufficiently responsive type of rational deliberation (as opposed to other

\textsuperscript{1194}And in this we already see some reason to think that conditions (1) and (2) will constitute a complete account of the freedom involved in action-responsibility only if we presuppose liberty of identification, i.e. in this case, that Mary could have identified with her motive, rejected it, or remained neutral as she did.

\textsuperscript{1195}Fischer, The Metaphysics of Free Will, pp.165-167. For example, a man who steals a book against his better judgment is responsible on Fischer’s model if there is some possible world in which the same mechanism operates and the reasons he has are sufficient to motivate him not to steal the book. But if he "would persist in stealing the book even if he knew that by so acting he would cause himself and his family to be killed, then the actual sequence mechanism would seem to be inconsistent with holding him morally responsible for his action" (p.167).
kinds of causes) but rather because we identify with the action — at least to the extent that it is in accord (rather than discord) with what we care about. We approve of this act in a way we could not if our only motive for doing it was the (real or perceived) impossibility of doing otherwise — for Frankfurt plausibly supposes that this is a motive with which persons can never identify.

Hence even if the desire we act on in the actual sequence was directly produced by some external stimulus or conditioning, rather than by a weakly-reasons responsive mechanism, we could be morally responsible for it if we identify with this desire (in the sense that the disposition to strong evaluations approving this desire is a lasting constitutive part of the cares we project, as explained in Chapter IV). There will be cases in which we are responsible on Frankfurt’s model, though we are not on Fischer’s, and cases in which we are responsible on Fischer’s model, though we are not on Frankfurt’s. Where these differences occur, it seems that the judgment of responsibility yielded by the identification-based model fits better with ordinary intuitions, or is more phenomenologically adequate. The phenomena of volitional identification are too important to be left out of an adequate actual-sequence account of action-responsibility.

VI.4 Frankfurt’s Theory of ‘Free Will’ is compatible with liberty of identification

Frankfurt distinguishes between freedom of action —the condition for moral responsibility for first-order acts— and freedom of the will: "When we ask whether a person’s will is free we are not asking whether he is in a position to translate his first-order desires into actions." While condition (1) of freedom of action concerns the relation of the will, to the behavior it tries to guide, freedom of

---

1196Note that it is identification as a phenomenon (however it is constituted) that plays a key role in Frankfurt’s account of action-responsibility; this account can thus remain convincing even if Frankfurt’s initial analysis of identification in his "Freedom of the Will" paper proved to be phenomenologically inadequate. When identification is understood not in terms of second-order desires, but instead in terms of strong evaluation, care, and projective motivation (as argued in Chapter IV), it becomes clearer why identification should have the role in constituting action-responsibility that Frankfurt’s account gives to it.

the will concerns the relation of volitions\textsubscript{2} to the desires\textsubscript{1} with which they identify the agent.

However, there is some ambiguity in Frankfurt’s account of this concept, since one can find two
distinct conceptions of free will in §III-IV of his "Freedom of the Will" paper. The first parallels his
criterion (1) for the freedom of action:

\begin{quote}
Analogously, then, the statement that a person enjoys freedom of the will means
(also roughly) that he is free to want what he wants to want. More precisely, it
means that he is free to will what he wants to will, or to have the will[\ldots] he
wants[\ldots].\footnote{Stern, Translator’s Introduction to Tugendhat’s \textit{Self-Consciousness and Self Determination}, p.xx.} \footnote{Ibid.}
\end{quote}

This analogy suggests that one’s will is free when one’s higher-order volition V\textsubscript{2} to act on a first-order desire D\textsubscript{1} is \textit{effective} in bringing it about that D\textsubscript{1} becomes one’s first-order will. In keeping
with the earlier analysis of condition (1) for freedom of action, I will label this FW\textsubscript{H}, for hypothetical
freedom of will: \textit{if} one acquires V\textsubscript{2}, then it will guide the first-order will to act on D\textsubscript{1}. As Paul Stern
says, an agent’s freedom of will in this sense "means that if his second-order volitions had been
otherwise, the first-order desires that moved him to action would have been correspondingly
different."\footnote{Frankfurt, "Freedom of the Will," p.20.} This says nothing about the agent’s actual capacity to form any given higher-order
volition (such as V\textsubscript{2}). Instead, it is simply "in securing the conformity of his [1st-order] will to his
second-order volitions...that a person exercises freedom of will" in this hypothetical sense.\footnote{Frankfurt, "Freedom of the Will," p.22. An example would be the addict who is not prevented from getting his drugs, but whose first-order will is still out of his control.} Thus, whatever higher-order volition is formed is not frustrated.

Frankfurt makes three important points about this kind of ‘freedom’ in willing. First, one can
have freedom of action without having FW\textsubscript{H}: "A person who is free to do what he wants may yet not
be in a position to have the will he wants."\footnote{Ibid.} Second, FW\textsubscript{H} is a stronger condition than simply
\textit{identifying} with D\textsubscript{1} through V\textsubscript{2}. We identify with our actual will\textsubscript{1} as long as it is the desire or
intention we will₂ to act on, but this agreement between our second-order volition and first-order will does not guarantee that the former guided or the latter in becoming our will₁.¹²⁰² Their agreement could simply be a fortunate coincidence. But hypothetical freedom of will requires something more: it requires that it was the second-order volition that secured the right first-order will, and so FW₁ is lacking both when they are discrepant (as in the unwilling addict) and when "their coincidence is not his own doing but only a happy chance"¹²⁰³ (as for the willing addict).¹²⁰⁴ A person with free will has the will₁ she has because of her higher-order will₂.

Frankfurt’s description of someone who lacks free will as "estranged from himself"¹²⁰⁵ because his second-order volitions are not satisfied is thus a somewhat simplified formulation. For example, consider someone who acts out of nastiness towards a relative when he wills₂ to act of out kindness instead. Because he is not moved by the sympathetic intention₁ with which he identifies, "he finds himself a helpless or passive bystander to the forces that move him."¹²⁰⁶ As such a case illustrates, the frustration of volitional identification entails lack of free will. But the inverse does not hold: as we saw, one can lack free will even when one’s second-order volitions are accidentally satisfied. For example, suppose that just before he meets his relative, our man finds a $100 bill; as a result, he is so happy that his urge to be cruel to his relative dissipates. He will surely still be somewhat

¹²⁰² Moreover, as we have seen, identification with D₁ can occur even when some other conflicting desire actually serves as our action-maxim, and D₁ remains only the counterfactual ideal we project for our first-order will.


¹²⁰⁴ This brings out again the point that guidance seems to involve the counterfactual conditions of hypothetical freedom: had the agent willed, otherwise, the first-order will would have been guided accordingly, if such a will were formed at all. This qualification is important, because it suggests that on Frankfurt’s conception at least, there can be cases where someone’s higher-order volitions identify them with their actual first-order will W₁, and yet fail to guide this W₁. Although the agreement between second-order and first-order will may be enough to make the action autonomous (see notes 38 & 39 below) it is not enough for free will in Frankfurt’s sense. Some of Frankfurt’s examples meet these conditions: e.g. someone could not have willed₁ otherwise, but identified nevertheless with their first-order will. In these cases, as I read ‘guidance,’ the necessity determining the first-order will shows that something other than the higher-order volition guided the first-order will.


¹²⁰⁶ Ibid.
disappointed with himself, since he knows that but for the lucky find, he would have been nasty to his relative in spite of himself.

This clarification, however, only strengthens Frankfurt’s argument for the third point, which is that \( FW_H \) is not required for action-responsibility: "it is not true that a person is morally responsible for what he has done only if his will was free when he did it."\(^{1207}\) Action-responsibility entails that he acted freely, but since this does not require that the will \( \omega \) on which he acted was freely willed \( \omega_2 \) or guided by his second-order volitions, \( FW_H \) is not required for him to be responsible for his bodily actions.\(^{1208}\)

Most commentators concentrate on hypothetical sense of free will in Frankfurt’s analysis. However, there is a second concept of free will implied in his "Freedom of the Will" essay that is much stronger than \( FW_H \) as I have outlined it. As we saw, hypothetical freedom of the higher-order will says nothing about one’s actual capacities for forming second-order volitions: it only requires that if on forms a volition \( \omega_2 \), this must guide one’s will \( \omega_1 \). But notice that on this definition, even a wanton who abstains from forming higher-order volition could enjoy \( FW_H \), as long as it is true of her that in any (relevantly similar) world in which she forms a volition \( \omega_2 \), this brings her will \( \omega_1 \) into line with it. But Frankfurt insists that a wanton can neither have nor lack freedom of will in his sense—it cannot even be an issue for the wanton.\(^{1209}\) Thus the sense in which wantons neither have nor lack

---


\(^{1208}\) Frankfurt, "Freedom of the Will," p.24. Frankfurt makes this point in a passage that is somewhat confusing in its wording: the assumption "that a person is morally responsible for what he has done does not entail that the person was in a position to have whatever will he wanted [\( FW_H \)]. This assumption does entail that the person did what he did freely, or that he did it of his own free will. It is a mistake, however, to believe that someone acts freely only when he is free to do whatever he wants or that he acts of his own free will only if his will is free [my italics]" (p.24). By acting ‘of his own free will’ Frankfurt means acting freely in his sense; the point he intends to make is that acting freely in this sense does not entail free willing or effective guidance by the higher-order will.

\(^{1209}\) Frankfurt, "Freedom of the Will," p.21. As we have already noted, of course, there is an ambiguity in Frankfurt’s concept of the wanton, since he does not distinguish between beings who have the capacity to form higher-order volitions but have not exercised this capacity (‘voluntary wantons’) and those who are essentially or always wanton. Essential wantons could never satisfy the antecedent of the condition for hypothetical (continued...)
free will is one that depends on using one’s capacity for form identifications. A later passage clarifies this point:

A person’s will is free only if he is free to have the will he wants. This means that, with regard to any of his [first-order] desires, he is free either to make that desire his [1st-order] will or to make some other desire his will instead. Whatever his will, then, the will of the person whose will is free could have been otherwise; he could have done otherwise than to constitute his will as he did.\textsuperscript{1210}

This much-overlooked passage confirms that for Frankfurt, free will requires both liberty or alternative possibilities of the first-order will,\textsuperscript{1211} and that our first-order will is formed in accordance with our higher-order volitions. As Fischer suggests, we can understand freedom of the will in Frankfurt’s account "to involve the power to make another first-order desire one’s will as a result of a second-order volition."\textsuperscript{1212} Thus in full freedom of the will, our liberty to make either first-order desire D\textsubscript{1}a or D\textsubscript{1}b our will\textsubscript{1} must derive from our prior capacity to form either the second-order volition V\textsubscript{2}a that identifies with D\textsubscript{1}a or the second-order volition V\textsubscript{2}b that identifies with D\textsubscript{1}b. Hence libertarian freedom of the will (FW\textsubscript{L}) as opposed to hypothetical freedom of the will requires that one’s higher-order volitions also could have been otherwise:

FW\textsubscript{L}:
\
(1) The person has the liberty to form any one out of a range of alternative possible higher-order volitions (or identifications); (2) When the agent chooses a unified second-order will V\textsubscript{2}, his or her will\textsubscript{1} (the motive on which the agent acts) is determined in accordance with V\textsubscript{2}, because V\textsubscript{2} guides the formation of the will\textsubscript{1}.

But like the weaker FW\textsubscript{H}, this stronger sense of freedom of the will is not required for action-

\textsuperscript{1209}(...continued)
freedom of the will, but voluntary wantons can.

\textsuperscript{1210}Frankfurt, "Freedom of the Will," p.24, my italics.

\textsuperscript{1211}Someone may object that the phrase ‘could have been otherwise’ is sometimes used in other ways. But Frankfurt typically uses it to designate a libertarian capacity to reach alternate possibilities—for example, in referring to PAP as the principle that responsibility entails that the agent could have acted otherwise than he did.

A person’s responsibility for his actions does not require that "the alternatives he opted against were available to him"—either in action or by an effective higher-order volition—according to Frankfurt. Thus although free will (FWL) involves a libertarian capacity, free will is not required for responsibility for one’s outward, bodily actions.

But notice that FWL involves two conditions that are in principle separable. Therefore, we could define another sense of free will, which I will call libertarian freedom of identification (or Fl), using only the first condition for FWL. The person who has Fl has the power or capacity to form alternative higher-order volitions out of some range R of possibilities permitted by his volitional nature (a range that could be wider or narrower). Fl, then, is a condition stronger than FWh but weaker than FWL: a person could have this liberty to identify himself volitionally with different possible motives for acting (or with different types of 1st-order intentions) even without the ability to actually determine his first-order will in accordance with his volitions. Thus for example, an addict enjoying such ‘liberty of identification’ is one who is free to be either a willing or an unwilling addict—although she cannot actually refrain from acting on the addictive desire, and hence cannot control her first-order will through her identifications. She has Fl but lacks FWL.

Liberty of identification or Fl, then, is freedom only of the higher-order will; it involves alternative possibilities of identification, but does not require that we have alternative possibilities of

---

1213 As Frankfurt says, "It is not true that a person is morally responsible for what he has done only if his will was free when he did it. He may be morally responsible for having done it even though his will was not free at all" (pp.23-24). While responsibility requires freedom of action, it does not require that the higher-order will causes the intention which (along with the movement) constitutes the action. In "Identification and Wholeheartedness," Frankfurt appeals to a slightly weaker condition than FWL, which he seems to equate with autonomy: an act is autonomous when the agent actively decides to identify with the desire or intention involved in his act, or make it a "desire that is incorporated into him," whether or not this identification has brought about that this desire became his first-order will (p.170). Yet it is not clear that an action needs to be autonomous in this sense for us to be morally or legally liable for it, given that Frankfurt seems to hold that wantons can sometimes be responsible for their acts, since they do not oppose the motives on which they act.

the first-order will. This suggests two very important points. First, Frankfurt’s idea that the will characteristic of persons is phenomenally hierarchical, or involves volitions with different levels of authority or internality for the agent, is in and of itself perfectly compatible with a libertarian understanding of the higher-order will. Second, and more importantly, even if Frankfurt-style counterexamples to PAP-type principles convinced us that we do not need to be able to will and act otherwise to be responsible for our action (and for the guiding intention that is a constitutive part of it), this leaves open the possibility that libertarian freedom of identification (FL) is required for responsibility. We would need a separate argument to show that FL is superfluous—an argument that neither Frankfurt nor anyone else has given.

On the contrary, I suggest that Frankfurt’s own examples of the willing and unwilling addict borrow much of their intuitive appeal for our inchoate understanding of moral responsibility from the fact that they would fit neatly into a theory that made liberty of identification indispensable for moral agents. For if we agree so readily that the willing addict is responsible for acting on her desire for heroin, even though she could not do otherwise, is this not (at least in part) because we are implicitly assuming that this addict could identify otherwise, or has it in her power to change her higher-order will and become an unwilling addict? Prima facie, at least, this explanation of our judgments in such cases seems very plausible. If it is wrong, and we can consider the willing addict responsible even if no other commitments of her inner self or cares that would conflict with addiction are possible for her, surely we need to be shown why this is possible.

VI.5. Moral Responsibility for Character Requires Liberty of Identification

As we saw in §4, Frankfurt’s account of responsibility for first-order acts, omissions, and

---

1215 Because FL doesn’t require that our higher-order volitions effectively guide our first-order will, its requirement of alternative possible higher-order volitions does not entail that we have matching alternative possibilities of first order will or intentional action.
consequences of actions makes reference to the phenomena of the agent’s intrasubjective identification with her own motives. But for this very reason, while it is immediately clear that Fischer’s reasons-responsive mechanism model implies an actual-sequence account of action-responsibility, it is not after all so clear that Frankfurt’s model is an actual sequence account. The model combining conditions (1) and (2) above is itself neutral on this question: it can only be an actual sequence account if identification itself does not have to involve a freedom or liberty definable only in terms of alternative possible identifications open to the agent. As we have seen in §§1-3 above, a phenomenology of responsibility for first-order actions can force this issue by showing that identification with the motives and intentions on which we act is relevant for action-responsibility. But this type of phenomenological analysis cannot by itself shed sufficient light on the origin of identification to resolve this question.

This historical ‘source-problem’ for identification can be resolved only through a phenomenology of our responsibility for character rather than first-order action. In this section, I will argue that despite Frankfurt’s own confidence that identification does not require libertarian origins, the phenomenology of our responsibility for ourselves as whole persons suggests that we must have libertarian freedom at the level of identification itself, whether or not we need the liberty to act otherwise for action-responsibility.

VI.5.1 Internal Compatibilism cannot avoid the Action / Character Distinction

In recent writings by several influential philosophers, one finds defenses of a popular form of compatibilism that we might call ‘internal compatibilism:’ it urges that the ‘freedom’ worth having, the freedom vital for moral responsibility and human dignity, cannot require freedom from determination altogether (since that would imply complete arbitrariness); rather, it can only require the right kind of ‘inward’ determination by the agent’s desires, concerns, and deliberations, and their
freedom from (direct) external manipulation. Nicholas Rescher summarizes one appealing form of this approach as follows:

To have free will is in the first instance to be in a position to act from motives, goals, and values that one has—to have one’s actions proceed from the "motivational" sectors of one’s thinking via one’s own decisions and choices. It is a matter of the "inner," thought-embedded determination of what one does... A free act is not one that is free of all determination but one free from determination of a particular sort, namely, from determiners that operate "externally" in total detachment from one’s wants, preferences, wishes, and the like. Its agent "could have done otherwise" all right, but only if the agent’s choices and decisions had been different—which itself could have happened only if his or her values, goals, and decisions had been different.1216

This is similar to Frankfurt’s account of free action, although he emphasized that guidance by reasons and agent-internal motivations is not coextensive with such factors causally producing the bodily behavior which (with its intention) constitutes an outward action.

Yet an important difficulty arises for this approach: we normally take ourselves and others to be responsible not only for ‘outward actions,’ in which some bodily performance is guided by an intention, but also for our patterns of desires, practical reasoning, plans of choice, and especially for the ultimate goals and values that (at least implicitly) inform these dispositions and give our character such unity as it may have. In short, we impute to responsible agents not only their actions, but also their practical character —and may even reserve the strongest sorts of evaluation for the latter.1217 But if actions are ‘free’ or morally imputable in virtue of their being guided by one’s "capacity to implement one’s own agendas—to act on the basis of one’s feelings, wishes, desires,


1217Some senses of "good" and "evil" can apply only to dispositions and tendencies of practical character, and only secondarily (if at all) to particular actions. For example, it is primarily a person’s character that is malicious because it is dominated by hatred or spite as its prime motivation, or charitable because it is unified through loving concerns for the well-being of others; individual actions are malicious or charitable only if they are guided by the kind of practical character to which these judgments primarily apply. In this respect, these judgments are unlike other ethical judgments, for example judgments that particular acts are morally wrong or right, because the latter may be made mainly (if not only) on the basis of the specific intention constitutive of this action and its relevant circumstances.
preferences, and the like," as Rescher puts it, or on the basis of these actions being in harmony (or at least not in discord) with our identifications, as Frankfurt claims, then on what basis is the guiding character or volitional identification itself said to be imputable? In the face of this issue, the ‘internal compatibilist’ cannot simply say that the criteria of responsibility for our practical character are analogous to those for outward action, that we are responsible for our dispositions of motivation and highest goals just because they were formed or guided by yet other processes of deliberation, desire, or valuation that are also internal to the agent in the relevant sense, without risking another vicious regress.

For his part, Rescher proposes a very simple answer to the question, "what if one’s agenda of choice and decision is itself set by factors beyond one’s control?" Then, he responds, if the character relative to which actions are judged free or unfree is itself determined by other agents, it is unfree, but if determined by "nature and nurture," or the general lawful fabric of the world, then it remains free and imputable. On this account, the freedom-condition for the imputability of practical character is inherently social or interpersonal: "Freedom pivots on an agent’s independence of other agents and their machinations—not on an independence of operative causes." Yet (although Rescher appears not to perceive this implication), the freedom-condition for the imputability of outward acts will not be specifically contrasted with determination by other persons in the same way: for on this account, a bodily action will fail to be imputable if it is directly determined by non-internal causes, be they the actions of Frankfurt’s devil/neurologist, or a random biochemical glitch in the brain. In other words, Rescher proposes to solve the regress problem for internalist compatibilism by accepting a disanalogy between the conditions of moral responsibility for actions

\[\text{[footnote]}\]

\[\text{[Footnote]}\]

\[\text{[Footnote]}\]

\[\text{[Footnote]}\]

\[\text{[Footnote]}\]

466
and character.

Yet the social criterion Rescher proposes for the latter case, which makes essential reference to our practical character not being engineered by other agents for their own purposes, remains highly questionable. On the face of it, it is hard to see why there is any distinction that should make a moral difference to our judgments about practical character between cases in which that character is directly imposed by another person (for example, by a deity), and cases in which it is ‘just the way we are,’ given our genetics and early childhood. Rescher leaves it unclear why we should think that ‘freedom’ of character is not as much a matter of the agent’s relation to himself as is freedom of action on the internal compatibilist account.1221

Like Frankfurt, then, Rescher wants to hold that "The freedom at issue is...freedom from external control—which is compatible with internally rooted determinism via an agent’s own motives and ‘states of mind.’"1222 But Frankfurt sees that referring responsibility for character to negative social freedom (lack of intervention by foreign agents) is dubious; he wants to retain intrapersonal conditions for character-responsibility as for action-responsibility. Thus in response to this question, Frankfurt recently said:

I believe that a person is morally responsible for his own character insofar as he takes responsibility for it. This does not entail that the person be causally responsible for having produced in himself the dispositions and other features of his psychic constitution that make up his character. A person can take responsibility for his character, or for the second-order volitions that comprise it regardless of whether he himself caused their formation. Regardless of how the various elements of his character were produced, he takes responsibility for them by identifying himself with them.1223

1221For example, consider an imaginary world in which there is only one agent. On most ethical theories, it seems more likely that in such a world, evaluative ethical judgments would still apply to her character than to her outward actions, since these cannot harm anyone else.

1222Rescher, p.147.

1223Frankfurt, "Reply to Davenport," Presented at the Eastern APA meeting, Atlanta, GA (December 1996), mss. p.2. Frankfurt then went on to sketch an analysis of ‘identifying with one’s character’ along the lines of his satisfaction-analysis.
This remark is fully in accord with Frankfurt’s earlier critique of Aristotle’s theory of responsibility for character, which he thinks focuses too much on whether we have caused or failed to cause the dispositional characteristics we typically exemplify in action:

The fundamental responsibility of an agent with respect to his own character is not a matter of whether it is as the effect of his own actions that the agent has certain dispositions to feel and to behave in various ways. That bears only on the question of whether the person is [causally] responsible for having these characteristics. The question of whether the person is [morally] responsible for his own character has to do with whether he has taken responsibility for his characteristics. It concerns whether the dispositions at issue, regardless of whether their existence is due to the person’s own initiative and causal agency or not, are characteristics with which he identifies and which he thus by his own will incorporates into himself as constitutive of what he is.\textsuperscript{1224}

"Character," so understood, is the set of characteristics with which we identify, and thus for which we are responsible, however they originated.

Although it moves in the right direction, however, I think this is still an incomplete conception of character. Although I have not had space to develop an adequate account of character in this dissertation, I have suggested that there is a significant division between (a) inward character, including our dynamic identifications or cares; and (b) outward character, including an array of first-order motives, dispositions to act on them, and other behavioral traits and psychic features that make up our interpersonal style. I agree with Frankfurt that we are responsible for our outward character insofar as we take responsibility for the features it includes through identifying with them, and that this does not require that we be the cause of these features (say, in the sense that they were literally produced by our internal process of practical deliberation and choice). Yet crucially, it remains to consider why we are responsible for the "identifications"\textsuperscript{1225} that make up our inward character,

\textsuperscript{1224}Frankfurt, "Identification and Wholeheartedness," reprinted in The Importance of What We Care About, pp.171-2.

\textsuperscript{1225}Or, as I have been doing, we may use Frankfurt's original phrase "higher-order volitions" here instead, as long as (in accordance with the argument in Chapter IV) we understand "order" in the qualitative sense of authority for the agent, rather than the merely iterative sense, and we understand "volition" in the projective (continued...)
through which, as Frankfurt says, we "take responsibility" for features or characteristics of our outward character. As we have seen, this responsibility cannot lie in some further state of identifying with these very identifications, if by this is meant a further state that might or not obtain. In order to avoid a regress, we have to regard identifications themselves as essentially inalienable, or (to put it in Sartrean fashion) as carrying the agent’s identification with them in their very being.

VI.5.2 The Problem with Frankfurt’s ‘Quick Argument’ that character-responsibility does not require liberty of identification

At this point, Frankfurt could revert to his claim that there is no question about responsibility for inward character as I have defined it, since the identification built into our identifications themselves means we are automatically responsible for our them in the same way as for outward characteristics. In other words, he could claim that just as identifying with features of our outward character makes us responsible for them, no matter how they originated, likewise we are responsible for our identifications or inward characteristics, no matter how they originated (even if these identifications were programmed by a devil or evil neurologist or caused by some other external force). So there is no need to postulate liberty in forming identifications to explain our responsibility for them. In §IV.4.1, I called this Frankfurt’s "quick argument" against the origin of identification objection (B1), and I pointed out that it is question-begging because we could not just assume that what makes identification qualitatively distinct from alienable desires does not have something to do with the kind of origin it must have.

Similarly, we can now add that this quick argument implies an analysis of responsibility for identifications that is circular. It is true that for a state to count as our identification (i.e. to have the right authoritative quality), it must satisfy two conditions:

1225(...continued)
rather than merely desiderative sense.
(A) We must be automatically identified with this state, and

(B) We must be responsible for it.

But no reason has been given to think that (A) explains (B) in this case. On the contrary, no states other than cares or identifications satisfy condition (A): to say that a state satisfies (A) is then just to say that it is an identification. But if we ask what conditions are satisfied to make us responsible for this state, which is part of what makes it our identification, it would be wholly uninformative to say: ‘we are responsible for this state because it is an identification.’ So then it is equally circular or trivial to cite a state’s satisfying condition (A) as the reason why we are responsible for it. We could just as well say that we are automatically identified with this state because we are responsible for it in the way that we are always responsible for our identifications, and this would fail to be an explanation in the same way.

VI.5.3 The initial presumption in favor of the Requirement of Liberty of Identification (RLI) as a principle for character-responsibility

As I suggested with reference to the willing addict case above, it seems initially plausible to believe that the following is a necessary condition on a person’s responsibility for his own inner character, or for his positive identifications:1226

**RLI: The Requirement of Liberty of Identification:**

A person is morally responsible for his lasting identifications and cares, or for the volitional character he has acquired, only if he could (at some time in the past) have formed a different inner character, or (equivalently) if he had the power to bring about different identifications and cares instead.

In deciding whether RLI is true, it is crucial to realize that everyday intuitions are not neutral on this issue: there is an initial presumption in favor of such a libertarian condition on responsibility for

---

I do not yet claim that this principle also applies to decisive identifications, or the highest-order will. Of course Frankfurt’s argument in favor of volitional necessities implies that it does not. But this issue requires separate consideration.
character. There are (at least) two important reasons for this.

First, the intuition that we have some liberty to determine our volitional character is rooted in the fact that people’s character is rarely if ever morally perfect, yet we hold them responsible for the imperfections in their character (however this is judged). In standard conceptions of virtue, for example, we say that someone ought to care about things he doesn’t, or that someone ought not to identify with some motives, she does identify with; for example, a willing addict ought at least to be an unwilling addict. And this implies the possibility of change. In instances of imperfection, then, the \textit{ought} \textit{→ can} rule always implies that we can (or at least could have in the past) identify otherwise than we in fact do.

Second, usually when we lack alternative possibilities, this is because our state is not a \textit{self-determined} one, but is brought about by forces that do not include our own values, deliberations, and choices. Usually the only assurance we seem to have that our state is self-determined is that we had alternatives in bringing it about. This concern about external control naturally extends to include values, identifications, care and inner character in general. Hence it is natural to presume that if our character is \textit{self-determined}, we must have (or have had) some liberty in determining it. This is the intuition underlying Fischer’s concerns that structural or "mesh" theories cannot assure us that identifications would be self-determined.

This is also why Susan Wolf argues that the Frankfurt/Watson "real self" view of persons as defined by a set of higher-order commitments or a system of "values" (as distinct from desires) does not square with our intuition that a person may have a corrupt valutional system or set of higher-order volitions through no fault of her own (say from a very distorted childhood).\textsuperscript{1227} Nothing in the Frankfurt/Watson model seems to rule out values or identifications with inappropriate origins, such as conditioning or brainwashing, yet we would naturally think that persons as excused from

\textsuperscript{1227}See Susan Wolf, \textit{Freedom Within Reason}, ch. 3, "The Real Self View."
responsibility for their character in such cases. This is precisely why there is an "origin of identification" problem (B1), as explained in Chapter IV. Requiring that identifications (or evaluative commitments, in Watson’s initial model) originate in libertarian choices might turn out not to resolve this problem, but at least the defeasible intuition that such a liberty of identification is necessary for us to be responsible for our volitional character is an immediate and prima facie plausible response to a set of judgments we are disposed to make about character-states with inappropriate or responsibility-undermining origins. No convincing phenomenology of character-responsibility can hide from these judgments or deny that we are disposed to make them, and this is the key reason why there is an initial presumption in favor of an RLI condition: it seems to explain these judgments.

As we saw, Fischer’s model is an attempt to address this problem of inappropriate or responsibility-defeating origins, and this is exactly why Fischer’s reasons-responsive mechanism conditions refer to the causal origins of our actions. But Fischer’s way of resolving this problem too quickly abandoned the relevance of identification for the conditions of responsibility. He does not see that with respect to psychic states on which identification is conferred, their causal history is not so important, as Frankfurt maintains. But on the other hand Frankfurt does not see (or won’t yet admit) that with respect to the volitional states that do the conferring, their causal origin is relevant, as Fischer sees. So as I suggested, Frankfurt and Fischer are each partially on the right track, but also partially on the wrong one, in a complementary way.

The same initial and defeasible plausibility attached to PAP as to RLI. But in the case of first-
order actions, our intuitions have been corrected on this point (so we are now supposing) by Frankfurt-style counterexamples to PAP: these have shown us that self-determination is a condition independent from liberty or alternative possibilities, since in some cases these conditions clearly come apart. But this does not by itself create any presumption that the same must be true for responsibility for volitional character. Instead, we start out on this issue with the same defeasible intuition that we must be at some liberty in forming our character if it is to count as self-determined in the sense sufficient for it to be morally imputable. If this intuition is to be modified as it was in the case of action-responsibility, we need the same kind of evidence that we had against PAP in order to correct it, since our intuitions about responsibility do not start out neutral on this issue of responsibility for character.

In response to the suggestion that RLI is a condition on responsibility for character even if PAP is not a condition on action-responsibility, Frankfurt argued that it is "implausible on the face of it" that moral responsibility for character should be structurally very different in its conditions from responsibility for first-order acts. In particular, Frankfurt thought it implausible that responsibility for character could require alternative possibilities if responsibility for actions does not.1229 But we have now seen why this disanalogy is not at all implausible: in short, if the reason that action-responsibility does not require alternative possibilities of action is that being in deep accord with the agent’s character is sufficient to make an action imputable, then obviously this same reason cannot apply to character itself all the way down, and therefore it would not be surprising if character-responsibility required alternative possibilities even if action-responsibility does not. If responsibility for action traces to responsibility for character, as Frankfurt’s own account of action-responsibility (with its implicit reference to identifications) suggests, then these two types of responsibility are not on the same plane: the former is subordinate to the latter. It is natural to expect

---
that items in such a hierarchical relation involve different conditions.

If action-responsibility is compatible with the lack of alternative possibilities at the level of first-order acts, Frankfurt’s model suggests that this is only because we can identify with (or at least refrain from alienating) motives, on which we would nevertheless have acted if we opposed them, or identify with actions, which we would have done anyway if we had not done them for reasons of our own. In other words, phenomenologically, we have a prima facie belief that action-responsibility requires the power to avoid what we do or to bring about alternative possible acts; we can find a place for responsibility in cases where we lack such alternatives only because we can see intuitively that identifying with an act makes us responsible for it whether or not we can do otherwise. But if it is only because of this that responsibility without alternative possibilities fits our considered judgments about cases, then we have no reason to think that the same should hold true at the level of identifications themselves. In other words, if we ask why various states of our moral psychology seem to have the property of being such that

(i) we can (sometimes) be responsible for them even though we lack any (robust) alternatives to them,

we see that this is only because such states have the more basic property of being such that

(ii) we do not necessarily identify with them, but we can identify with them (our identifying with them is contingent).

The phenomenology of the cases which cast doubt on PAP for action-responsibility at the same time suggests that property (i) supervenes on property (ii). But states of identification themselves cannot have this more basic property (ii), since they instead have the property of being such that

(iii) we cannot alienate them (they are essentially or automatically identified with us).

Thus, lacking property (ii), identifications cannot have property (i). There is no reason we have so far encountered in the phenomenology of cases suggesting that some states have property (i) to think that any state besides those with property (ii) could have property (i); that is, no evidence has been
given that any state with property (iii) could also have property (i). Let us consider, then, how we could go about giving evidence that states of identification, which all have property (iii), also have property (i); we will try to construct examples to show this.

VI.5.4. The Impossibility of True Frankfurt-Counterexamples to RLI

(I) Suppose we draw up a case as follows. After some doubt about its effects on his relation to his own family, Romeo identifies with his desire to marry Juliet. And his identification issues from (or is caused by) a weakly reasons-responsive mechanism: there are possible worlds in which the same mechanism operates but Romeo alienates this desire (e.g. worlds in which he knows that this desire will lead not only to some family feuding, but to the destruction of the whole human race, or something equally outlandish). So his identification was sane, even if ill-fated, and on these grounds, he seems to be responsible for it. But unbeknownst to Romeo, if he had shown any inclination not to identify with the desire to be with Juliet always, then an evil magician (who likes tragic endings) would have cast a spell that made him identify with this desire and pursue not only its end (being with Juliet) but also the cultivation of the desire itself as a rich and beautiful thing. So the worlds in which Romeo alienates this desire are not accessible to him: Romeo cannot actualize any of these worlds. But surely he is still responsible for his identification even though he could not have identified otherwise.

This is would be a Fischer-style example of a case in which Romeo’s identification had both properties (i) and (iii). It depends of course on the assumption that a state’s being produced by a weakly reasons-responsive mechanism is sufficient for the agent’s being responsible for it. But unfortunately, in §1 above we found this account of responsibility wanting even as an explanation of responsibility for first-order actions (plus omissions and consequences), and therefore we have no reason to believe it can be extended to responsibility for inner character. Indeed, we were forced to
consider the conditions of responsibility for volitional identifications just because we concluded, contra Fischer, that these states played an indispensable role in explaining action-responsibility, and in particular in explaining how it is that other types of states can have property (i). Thus it would be inconsistent for someone who, on these grounds, favors Frankfurt’s account of the conditions for imputable or free action, and hence pursues the origin of identification question, now to agree that such a Fischer-style example proves that we can be responsible for identifications while lacking alternatives to them. That Romeo’s identification with his desire to be with Juliet issues from a reasons-responsible mechanism cannot (by itself) be what explains his responsibility for it.

(II) Suppose instead that we try to construct a counterexample to RLI that exactly parallels the kinds of fail-safe counterexamples Frankfurt devised against PAP. Screwtape is a nefarious devil1230 who watches Kevin, a person struggling with a minor character-flaw: in this case, Kevin’s desire to seek thrills through dangerous stunts (e.g. jumping a motorbike over cars, etc.). The question of whether this is foolhardy desire should be acted on has been repeatedly posed to him by friends and family. Yet Kevin finds ‘on his own’ that he cares deeply about experiencing the thrill of danger; on his own, he decides to evaluate his desire positively, and he builds up a disposition so to evaluate it. Eventually, he projects for himself the end of maintaining and cultivating his desire, which is to say that Kevin starts to identify with his dangerous desire. But let us suppose that, had it appeared that he was not going to identify with the desire, but rather to alienate it (and Screwtape is a very good judge of these things, almost never wrong), Screwtape would have correctly predicted which way Kevin was leaning, and would have intervened: whatever tendencies and evaluative dispositions constitute identification with the foolhardy desire for dangerous thrill, Screwtape would directly have caused these states to appear in Kevin. In other words, he would have programmed Kevin to identify with his desire.

Is the scenario this case describes conceptually or phenomenologically possible? If so, it would suggest that we can be morally responsible for our identifications even when we could not have identified otherwise. But if we hold with Frankfurt that identification refers to the authority that makes us responsible, and hence that we are responsible for any state with which we identify, then notice that although the Screwtape-Kevin case seems to parallel Frankfurt’s counterexamples to PAP, this case is unlike those earlier counterexamples in one crucial respect: in the counterfactual sequence, Kevin will still be responsible for his identification, whereas in the counterfactual sequences of the earlier counterexamples, Jones was not responsible for his action. The earlier counterexamples were persuasive because they could accommodate our abiding intuition that if we are directly forced to do something, or manipulated into doing it (or into forming an intention to act that way), then we are not responsible for that action — the very intuition that underlies the initial plausibility of RLI, as I explained. Thus if Black had intervened and made Jones turn left at the fork, Jones would not be responsible for turning left. As it is, though, Jones turns left for his own reasons: this means (as I have argued) that he acts on motivations with which he is at least tacitly identified. So these earlier counterexamples were able to eliminate Jones’s possibility of acting otherwise than he does in the actual sequence, and yet leave him responsible in that sequence while he is not responsible in the counterfactual sequence, because of the difference between these two sequences: he is manipulated in the latter, but not in the former. And as we saw in responding to David Copp’s "Madeleine" case (above), it is crucial that the agent can avoid blame in the counterfactual sequence.

But the putative counterexample we have now devised against RLI does not meet this requirement for persuasiveness. For by hypothesis, Screwtape is able to directly cause an identification to appear in Kevin, that is, a motivational state qualitatively different than others in part because it automatically includes Kevin’s identifying with it. In this counterfactual sequence, then, Kevin identifies with his identification and (given the meaning of identification) Kevin thus takes
responsibility for his identification and it counts as self-determined. Yet this fails to accommodate
the intuition that we are not responsible for something directly produced in us by forces that do not
work through our own agency. To remain plausible, as we saw, the Frankfurt-cases have to allow us
the possibility of avoiding blame. Imagine that we revised the earlier counterexample to PAP such
that Black directly causes Jones to decide to turn left on his own. Would this type of example
warrant the rejection of PAP as strongly as the original version of the example did? Clearly not; the
revised example would be far less persuasive, since the whole question is whether determining
someone to do something for their own reasons even makes any sense, or whether it is a
contradiction. Examples of this type, then, do not work with our initial intuitions to modify them in
a compatibilist direction; rather, they depend directly and obviously for their very possibility on the
claim that a person’s self-determined states can be determined by prior causes not internal to the
person’s agency. Unlike the earlier counterexamples against PAP then, the putative counterexample
we have devised against RLI fails to do any significant phenomenological work against our initial
defeasible intuitions in favor of RLI.

Alternatively, we could say that what the Screwtape-Kevin case shows is that, although we can be
responsible for identifications when we cannot identify otherwise, we are only responsible for our
identifications when we form them ‘on our own.’ On this reading, Kevin would not be responsible
for his identification in the counterfactual sequence where Screwtape induces it. This seems to
restore the conditions for plausibility, but at a heavy price. For then it seems that in the
counterfactual sequence, Screwtape does not actually induce an identification with the dangerous
desire for thrills after all: for to identify with a desire just is to take responsibility for it, and there is
no question that ‘taking responsibility’ is itself something for which we automatically take
responsibility. If it were possible for Screwtape to induce an identification by manipulation in the
counterfactual sequence and for Kevin not to be responsible for it because it was artificially induced,
then the Qualitative Condition would be violated.

In sum, we are thus faced with a very interesting trilemma concerning the Screwtape-Kevin case. What Screwtape induces in the Kevin in the alternative sequence is either an identification, or it is not. (a) If it is not, then the case as described (the case that would be a counterexample to RLI) is not conceptually possible after all, since Kevin can avoid identifying with his dangerous desire for thrills: in the counterfactual sequence he does not identify with it. If instead, what is induced in the alternative sequence is literal identification with the foolhardy desire, as the case demands, then either Kevin is responsible for the identification in the alternative sequence, or he is not. (b) If he is not responsible for it, then the Qualitative Condition is violated, which is contrary to the phenomenology of identification (Chap. IV-V), and thus phenomenologically impossible. (c) But if he is responsible for it, as the qualitative nature of identification demands, then as just argued, the case is unlike standard Frankfurt-counterexamples to PAP-type principles, and it loses its phenomenological appeal.

This trilemma shows that the Screwtape-Kevin case is either conceptually impossible, phenomenologically impossible, or phenomenologically implausible, depending on how we construe it. It is thus of little value to opponents of RLI. As a result, Frankfurt may say that

The fact that a person takes responsibility for something does not entail that the person could have refrained from taking responsibility for it or that he could have taken responsibility for something else instead. Whatever mental acts are required to accomplish the acceptance in which identification consists, a person may be morally responsible for performing those acts even if he could not have done other than to perform them. There is no reason why it should be any more essential for an agent to have alternatives to these [second-order or higher-authority] acts, in order to be morally responsible for performing them, than it is essential for morally responsible agents to have alternatives to acts of other sorts.1231

But no phenomenologically sound evidence1232 has yet been found for the claim made in the first two

\[ \text{\footnotesize 1231} \text{Frankfurt, "Reply to Davenport," p.3-4.} \\
\text{\footnotesize 1232} \text{There may be other kinds of evidence, e.g. the conceptual evidence afforded by arguments that RFI leads} \\
\text{\footnotesize (continued...)} \\
\]
sentences of this quotation, whereas evidence has been presented against the claim that alternative possible first-order acts must be accessible to the agent who is morally responsible for her first-order act. This difference in available evidence suggests that parity-claim made in the last sentence quoted here is false.

(III) A genuine Frankfurt-style example of an identification that has property (i) in addition to (iii) would have to go rather differently. To be structurally like the Frankfurt-style counterexamples to PAP with which we are familiar, it would have to be a case in which

(a) we are responsible for our identification in the actual sequence because we identify with it;

(b) in all counterfactual sequences where we begin or try to identify otherwise, we are blocked from forming alternative identifications by fail-safe mechanisms or interveners;

(c) in these counterfactual sequences, we do not identify automatically with our imposed identifications in the very act of acquiring them, and thus we are not responsible for the identification imposed on us;

But since all identifications have property (iii), there are no worlds in which we do not identify with our identification. It is not that our access to this possibility is blocked; it doesn’t exist. So no Frankfurt-style example of responsibility for identification can satisfy condition (c).

Well then, can we not simply accept that the Frankfurt-style examples for identification-responsibility must differ structurally from those for action-responsibility in this one respect, but still insist that such examples satisfy conditions (a) and (b)? Certainly allegedly possible examples like this can be proposed, but as I have argued, they will not do any significant work. Because condition (a) is necessarily rather than contingently satisfied, it does not add in anything that is missing in the counterfactual sequences that could explain why we are responsible for our identification even though we could not identify otherwise, and Jones’s deciding to turn left on his own in the actual

---

1232 (...continued)

to the aporia of arbitrariness, but this is a different matter, to be addressed separately.
sequence does in the counterexample to PAP.

We are forced to see that the kind of identification that is \textit{built into} states of identification themselves is very different than the kind of identification that is \textit{conferred on} other states with which we do not automatically identify (and that apparently allows us to be responsible for these states on which it is conferred even when we cannot avoid these states). And this undermines any intuition we might have that responsibility should naturally work the same way at the level of identifications themselves. This disanalogy arises just because of the \textit{qualitative difference} between the levels of conferred and immediately built-in identification: the intuitions that (at least seem to) tell us the former is sufficient for us to be responsible for states on which it is conferred, are not intuitions about the latter at all. There is thus no intuitive support for the supposition that, if we are always responsible for our identifications, this is merely because of the identification with identifications that is automatically built into them. And without a reason to think otherwise, our initial defeasible intuition that alternative possibilities are necessary for responsibility remains in full force at the level of identifications themselves.

This defense of the RLI principle for character-responsibility works by drawing out the implications of a fact noted earlier. If we act on a given motive, and we identify with that motive (actively and positively, or passively and tacitly), then the identification conferred on the motive makes us responsible for acting on it, even if we could not have avoided acting on that motive. I noted that conferred identification produces such exceptions to PAP in a fashion analogous to the way a tracing principle for the transfer of responsibility allows for similar exceptions. Such tracing principles are intuitively appealing because they trace our responsibility for actions or states of affairs we cannot avoid to past choices we could have avoided at the time. Similarly, Frankfurt’s principle for responsibility traces our responsibility for actions we could not avoid to our identification with them, and as we now can see, just like a tracing principle, the phenomenological
appeal of this principle depends on the assumption that we could have avoided conferring identification on the motive with which we did identify.

VI.6 An Existentialist Response to the Challenge of Volitional Necessities

VI.6.1 Volitional Inertia vs. Volitional Necessity in Frankfurt’s Examples

At this point, I would expect Frankfurt to say that even if no convincing counterfactual-intervener or fail-safe overdetermination counterexamples to RLI can be devised, there is a different type of evidence against RLI. For Frankfurt has argued in a number of essays now that persons can find themselves necessarily identified with certain motives, or (alternatively put) unable not to care about being a certain sort of person (in the sense of outward character). I already mentioned Frankfurt’s examples of Lord Fawn (V.2.1), Luther (IV.5.2), and his notion that genuine love is a state of volitionally necessary identification with the desire to promote the beloved’s well-being (IV.5.2):

"The lover cannot help being selflessly devoted to his beloved. In this respect he is not free. On the contrary, he is in the very nature of the case captivated by his beloved and by his love. The will of the lover is rigorously constrained." But this does not mean that love is not self-determined; as in the other cases, this is an internal constraint, a necessity of identification itself, that seems irresistible to us not in sheer motivational power but rather in authority.

Whether we are convinced by the description of these particular examples, however, it does seem plausible that there can be cases in which our identification is so steadfast that we cannot identify otherwise, and yet this only shows how essential the identification at issue is to the inner character for which we are responsible. Frankfurt mentions "a mother who reaches the conclusion, after


conscientious deliberation, that it would be best for her to give up her child for adoption," but who finds when the moment arrives that she cannot go through with it. In this case, we have to imagine that this is not because of any fears that compel her although she does not identify with them; instead, it is precisely the strength of her identification with her concerns for her child that makes her horror at giving the child up insurmountable. We could find or make up many similar examples. Consider a person who is so unmovably committed to pacifism and non-violence that she will not sell a small amount of her country’s resources to a warlike power even though this means the complete annihilation of her people. Such identifications represent the most resolute stands we take: if we are not responsible for them, we cannot be responsible for any part of our character. And although such particular and principled commitments are characteristic of agents with a developed sense of self and well-organized characters, some limits on what we could bring ourselves to will are surely vital for anyone sane enough to count as an agent: as Eleonore Stump likes to say, it is volitionally impossible for a normal adult to consider accepting a nickel to cut up her children. This is not simply because we are averse to harming our children for almost any reason (even meritorious ones); as Frankfurt says, we also endorse or identify with such an aversion, and this is part of why it constrains our conduct so effectively.

Do examples of this sort show that the RLI principle for character-responsibility is false? The existentialist defender of RLI has several possible responses. Although a fully adequate reply would require developing a complete existential account of the highest-order will or decisive identification—a task that remains beyond the scope of this dissertation—I will briefly sketch how these different

---

1236 The original Star Trek episode "Mirror Mirror" includes a race called the Halcans whose leaders take this position against an overwhelming threat from 'twin-Enterprise.' Frankfurt also mentions military officers who found themselves incapable of carrying out orders to launch nuclear missiles when they thought the orders were for real ("Rationality and the Unthinkable," p.182).
1237 Frankfurt, "Rationality and the Unthinkable," p.182.
lines of response should fit together when fleshed out.

First, it is possible to confuse volitionally necessary identification with the ‘staying power’ that all states of volitional care must have to count as cares that establish our identifications at all.

Frankfurt’s analysis of love may fall into this confusion. In an important passage, Frankfurt describes a qualitative contrast between states of love and elementary psychic states that do not intrinsically carry the agent’s identification or authority with them:

The volitional attitudes that a person maintains towards his own elementary motivational tendencies are entirely up to him. Passions such as jealousy and craving merely provide him with psychic raw material, as it were, out of which he must design and fashion the character and structure of his will. They themselves do not essentially include any affirmative or negative volitional attitudes towards the motivational tendencies in which they consist....

However imposing or intense the motivational power that the passions mobilise may be, the passions have no inherent motivational authority. In fact, the passions do not really make any claims on us at all....

Love is different. The fact that someone loves something does not merely present him with primary volitional raw material. Love is not an elementary psychic datum, which in itself implies no evaluative or practical attitude on the part of the lover towards its motivational tendency. To be sure, a person may regret loving what he loves....No doubt there are many possible varieties of volitional complexity and ambivalence. But since love is itself a configuration of the will, it cannot be true of a person who does genuinely love something that his love is entirely involuntary.

This does not mean, of course, that a person necessarily assigns the very highest priority to serving the interests of whatever he loves....The fact that a person loves something does imply, however, that he cannot help caring about its interests and that their importance to him is among the considerations by which he cannot help wanting his choices and his conduct to be guided...It is an element of his established volitional nature, and hence of his identity as a person.  

This analysis rightly recognizes the qualitative difference between identifications or cares and other motivational states. It is an essential property of our loves that we identify with them, and this does not hold for our passions. All states of care are necessarily (de re) identified with. But this does not imply that it is an essential property of the agent to have a given love. What follows is only that it is necessary (de dicto) that if an agent loves or cares about something, then the agent identifies with so

---

loving or caring.

However, in this passage Frankfurt also has in mind another connection between this essential property of a care or love and its agent: the authority essential to states of love implies that the agent cannot simply reject this love ‘at will.’ As Frankfurt says, "We are certainly not free to decide ‘at our own liking’ what to love or what love requires of us." And in this respect, a loving agent "cannot directly affect his will by a mere act of will"1239 or by the decision to form contrary intentions (as we saw in IV.4.3). But this staying power or temporal inertia that is part of the essential quality of cares or states of identification does not mean that it is ‘volitionally impossible’ for the agent to identify otherwise in the broader sense in which it is impossible, e.g., for Luther to retract his theses. As Frankfurt says, although the process of an agent’s "coming to care about certain things" may not be "wholly under his voluntary control, it is nonetheless often possible for him to affect them,"1240 in particular by attending to different grounds for projecting motives, or reasons to care one way rather than another.

For the same reason, we can also be ambiguous about what to care about, or divided in our loves. By contrast, the kind of volitional necessities that apply to "ideals" we cannot bring ourselves to violate "constrain us from betraying the things we care about most and with which, accordingly, we are most closely identified."1241 They represent decisive identification or maximally authoritative commitments, and as the possibility of ambiguity about love suggests, not every instance of care need be this strong.1242 There is, accordingly, a difference (in degree or kind) between the sense in which every identification is dispositional and thus resistant to change through a single decision, and

1239Ibid, p.441.
1240Frankfurt, "The importance of what we care about," p.91.
1241Ibid, p.91, my italics.
1242Frankfurt recognizes this when he writes: "About certain things that are important to him, a person may care so much, or in such a way, that he is subject to a kind of necessity" see "On the Necessity of Ideals," p.20). This implies that volitional necessity is a higher form of caring or loving.
the sense in which a volitional necessity cannot be altered "merely at will," although Frankfurt tends to run these senses together.

VI.6.2 A Sketch of Constrained Liberty in the Highest-Order Will

This important qualification suggests a second point. A defender of RLI could argue that whereas volitional necessities constitute decisive identifications, the RLI principle is only meant to apply identifications of that middle order of authority that still allows conflict and ambiguity. But the existentialist will not take this approach, because (as we saw in discussing Tugendhat —§V.3) the existentialist maintains that even at the level of our ultimate commitments —our "ground projects" in Williams’s sense or our "original projects" in Sartre’s sense— the agent can over time bring about non-random though unavoidably radical changes. And this liberty in the highest-order will is essential to our responsibility for our deepest selves on the existentialist account; even our innermost identity is not simply given to us as a predefined personal essence or individual nature, but self-fashioned by the ‘I’ in time.

This can be true even though the ‘I’ that fashions its maximally inward self is never a bare ‘I’ without any prior volitional content or orientation, or never fashions itself ‘from nothing.’ The existentialist’s third point in response to Frankfurt is that liberty in the highest-order will can be a constrained liberty: it is compatible with (and even requires) that relative to any short duration of time, there will always be substantive limits on what is volitionally possible for us to will, not only in our intentions, and our identifications, (which are guided by our highest-order projections), but even more so at the level of our decisive or unified identifications. This fits with the picture of the person as a hierarchy of three nested ‘spheres’ or horizons of volitional possibilities of the three different relevant orders of identification (IV.6.1): at the highest-order level, the range of identifications that

are *volitionally possible* for him are very tightly constrained by their relation to the particular identifications, the agent already factically has. An absolute pacifist cannot just suddenly take up savage violence as the ultimate end of her life. But this does *not* mean that the scope or range of these volitional possibilities cannot shift over time, or that this shift cannot (at least in part) be affected by the agent. Persons are agents who live in time, and constantly face new circumstances, in light of which more remote volitional possibilities may become slightly more actionable, while other more central ones may shift to slightly less dominant positions in our practical understanding. Moreover, how personal agents react to new circumstances, and new self-interpretations that may accompany them, may even having the effect of bringing previously unthinkable options within the horizon of volitional possibilities, even though at first usually only as very remote possible actions or motives for acting. Through such a process, and sometimes through dramatic upheavals in moments of crisis, we can change our decisive identifications, and even alter what is volitionally necessary for us, developing new volitional necessities and losing old ones.

In several passages, Frankfurt himself concedes that volitional necessities may not only change, but that the agent can at least indirectly guide this alteration. He allows that such necessity is "to a certain extent self-imposed," since the person maintains his care and intentionally avoids "being guided in what he does by any forces other than those by which he most deeply wants to be

---

1244 Heidegger appears to have something like this process in mind in his account of the simultaneously factical and transcendent (mood-conditioned yet futural-projective) character of our existential "understanding" of what we can do and what relations we can take up to entities within-the-world. See *Being and Time*, ¶31, H144, p.183: "Possibility as an *existentiale*, does not signify a free-floating potentiality-for-Being in the sense of the ‘liberty of indifference’ (*libertas indifferentiae*). In every case Dasein, as essentially having a state-of-mind, has already got itself into definite possibilities...Dasein is a being-possible which has been delivered over to itself—*thrown possibility* through and through."

1245 See Sartre’s description of radical moments of conversion in *Being and Nothingness*, p.612: "One may recall the *instant* at which Gide’s Philoctetes casts off his hate, his fundamental project, his reason for being, and his being. One may recall the *instant* when Raskolnikoff decides to give himself up. These extraordinary and marvelous instants when the prior project collapses into the past in the light of a new project which rises on its ruins and which yet exists only in outline, in which humiliation, anguish, joy, [and] hope are delicately blended, in which we let go in order to grasp and grasp in order to let go—these have often appeared to furnish the clearest and most moving image of our freedom."
guided." This implies that we could have done otherwise than to maintain his care in this way, and that its necessity for him is constituted in part by the fact that he does not move to alter it. The fact that he does not, however, is not determined simply "by performing a voluntary act" of decision. The existentialist will account for this by saying that, like the care it maintains, this project of retaining the same dispositions of inner character is itself dispositional. But this does not mean that this third or highest-order projection of identifications is something the agent cannot consider changing or bring to the point of fundamental alteration over time. Even less than ordinary identifications is it "subject to his immediate voluntary control," but that does not mean it is beyond his regulation entirely. As Frankfurt says in a later essay,

...what is unthinkable for a person may vary from one time to another. The necessities of the will are not necessarily or always permanent. They are subject to change, according to changes in the contingent circumstances from which they derive. It is even possible for a person to desire that the limits of his will should be different from what they are, and to attempt deliberately to change them....Needless to say, it would be out of the question for him to change himself in any such way by a sheer act of will [or single decision]....On the other hand, it is surely open to someone for whom an action is unthinkable to try by other means, less direct than the exercise of willpower alone, to alter his own will in such a way that the action becomes thinkable for him.

The existentialist will explain this as follows. Within the pattern of dispositions, judgments and attitudes making up an identification, there will always be alterations over time, even though they do not destroy the pattern as a whole. On what grounds (which can shift with new experience) and in just which ways the agent is disposed to continue projecting the motivations involved in this pattern will determine whether there is an inner tendency towards or away from a cumulative set of changes

1246 Frankfurt, "The importance of what we care about," p.87.
1248 Ibid, p.89.
1249 Frankfurt, "Rationality and the unthinkable," p.187. Frankfurt mentions the imaginary example of a man who cannot identify with a desire to eat human flesh, no matter how hungry he is. He imagines that this man could try to change this, or bring himself to the point where he could identify with acting on this cannibalistic intention. A passage very similar to the one quoted here is found in "On the Necessity of Ideals," p.21.
in motivation sufficient to break the pattern as a whole and establish a new one in its place. It is this threshold effect which allows apparently radical changes in fundamental commitments to be self-directed, being preceded by tacit and underlying shifts in which the agent becomes less ‘comfortable’ with her ultimate motivations, or ground projects.

Fourth, the existentialist can also accept (and may even require) that this type of development—the way that different sorts of action and choice affect our character, altering our dispositions and even what is volitionally possible for us—may itself be governed by general indeterministic laws, which explain why only certain conceivable histories of character-development are really possible for us. To clarify what I have in mind, consider this loose topological analogy: such spiritual laws would explain why certain paths through the existential ‘space’ defined by time and character as axes are like deep groves or courses into which our individual developments have a tendency to run, so that getting from one riverbed to another in our existential space requires an uphill climb over divides—divides that grow higher and steeper the further along we are in time—whereas ‘going with the flow’ requires constantly less and less initiative once our character patterns are following along in a well-defined groove.

VI.6.3. The Phenomenological Problem of Individual Volitional Essences

Although the details of this account still need to be worked out, the central point for the existentialist is that none of the substantive content of an individual’s highest-order identifications—as opposed to the spiritual laws or dynamic forms of character-development which are similar for everyone—is necessary in the strong sense of being absolutely unchangeable through any process guided by the agent himself. For suppose there were volitional necessities that are necessarily fixed for an individual throughout his entire life, as opposed simply to being contingently kept up by the individual although at some points he could have directed his development in a way that would
eventually have changed them: call these absolute volitional necessities. Such necessities of character would constitute place some aspects of an agent’s character completely beyond even his longest-term control and constitute the kind of personal essence that existentialism is committed to denying, since who we are as an individual cannot simply be innate in us.

In places, however, Frankfurt says that some of our volitional necessities are absolute and do constitute a personal essence in this sense. For example, he writes:

A person for whom an action is unthinkable may be in a position to alter his will by means less immediate and direct than the exercise of willpower alone. However, undertaking to make the unthinkable thinkable might itself be something the person cannot bring himself to do....To be sure, changes in circumstances might bring about a change in his will. But he cannot change his will by an deliberate effort of his own. He is subject to a necessity that, in this sense, defines an absolute limit.  

Frankfurt says that volitional necessities of this kind constitute one’s "essence as a volitional creature;" they are constitutive of the agent’s "nature or essence as the person he is." Similarly, he writes that "The essence of a person...is a matter of the contingent volitional necessities by which the will of the person is a matter of fact constrained....These constraints cannot be determined by conceptual or logical analysis. They are substantive, rather than merely formal." In other words, they define the individual, rather than personhood in general.

This analysis returns us to a Spinozistic metaphysical naturalism about personhood at the expense of freedom. On the picture it presents, for each person, there is a set of absolute identifications that supply the metaphysical basis of the agent herself (in the same way as Stump equated the agent with a particular mind). Behaviors only count as her actions because they have an intentional explanation within the horizons set by these limits. Initiatives can be said to stem from her only because they

---

1251Frankfurt, "Rationality and the Unthinkable," p.188.
1252Frankfurt, "Autonomy, Necessity, and Love," p.443. Notice that this definition seems somewhat less exact, since it obscures the previous distinction between the ordinary and absolute levels of volitional necessity.
have the authority of these identifications—which literally are ‘her’—behind them. Frankfurt thus proposes a *volitional reification of the person.*

This reification is disturbing for two interrelated reasons. First, if external forces or changes of natural fortune *could* change the absolute volitional necessities that define a person, in what sense could we say that it is the *same* person before and after the change? Continuity of identity through such radical changes seems to require that the earlier person had a hand or role in bringing about such changes, as Aristotle thought. Second, if the agent cannot affect or guide the absolute identifications that define him, his most inward character or moral identity is simply *created* for him by fate, or biological processes, or God, or whatever. In such a case, our *prima facie* intuition would be to say that the agent is not responsible for this absolute core of his volitional identity, and hence not for any of the authority it confers on other aspects of his character either. We would need examples and cases to show us that a person can be *responsible* for absolute identifications that we simply built into him from the beginning, or by virtue of which there is a ‘him’ at all.

But the cases Frankfurt has presented do not provide phenomenological evidence of identifications that were always absolutely necessary to the person, or became so through processes out of his control. In any plausible case, I suspect we will find that, although the volitional necessity seems *at a given point* in a person’s life to be unchangeable, this is a result of a prior pattern of identifications that were not themselves necessary for the agent in the same sense. In other words, the existentialist can appeal to a tracing principle: it is possible for an agent to be accountable for

1253 Though it is less rationalist, this reification is in certain respects like the one effected by Kant in his late doctrine of the ‘highest maxim’ that functions as a noumenal and atemporal determinant of a person’s moral character or practical identity.

1254 See *Nicomachean Ethics*, Book III.5. Williams’s analysis in his famous essay, "Persons, Character, and Morality" supports this point as well.

1255 Frankfurt seems fully prepared to accept this. He writes: "We can only be what nature and life make us, and that is not so readily up to us" (see "The Faintest Passion," p.10). Yet note that Frankfurt still seems unwilling here to say outright that the most important part of who we are is determined by nature and life in a way not "up to us" *at all.*
what is now a fixed aspect of his innermost character or highest identification, only because he could have avoided it at sometime in the past. This historical modification of the RLI principle implies of course that there are no volitional necessities that are absolutely unchangeable or fixed throughout a person’s life, which would define his core identity—who he is in his life as a whole—before he is born or has lived as an agent at all.

VI.6.4 Between Frankfurt and Sartre: Neither Volitional Essence nor Absolute Groundless Freedom

There is no phenomenological evidence against this existential principle ruling out absolute volitional identities. Instead, Frankfurt presents a purely conceptual argument to show that some of the volitional necessities we experience must be absolute, since on transcendental grounds, agency requires some basis in a fixed volitional essence. In Frankfurt’s view, the alternative is a kind of liberty in the highest-order will that would lead to a picture of the person as a bare or factically unencumbered ‘I’ whose choices could only be arbitrary: "an excess of freedom gives rise to a diminution, or even to a dissolution, of the reality of the self." On the contrary, he holds that the nature of a person’s will "is not necessarily altogether formal and contentless, having no inherent proclivities of its own." Each individual will have his or her own unique volitional shape at the innermost level of agency.

Frankfurt presents his most developed version of this argument in his essay, "On the Necessity of Ideals." He begins with the familiar distinction between self-determination and liberty. The latter ideal is advanced by the "steady expansion of the range of options from which people can select" in life. But there is also what he calls

1256 Frankfurt, "Rationality and the unthinkable," p.179.
1257 Ibid, p.190.
the ideal of individuality, construed in terms of the development of a distinctive and robust sense of personal identity. To the extent that people find this ideal compelling, they endeavor to cultivate their own personal characteristics and styles and to decide autonomously how to live and what to do.\textsuperscript{1259}

"Individuality" in this sense seems to be a combination of volitional self-unification through decisive identification or wholeheartedness, and an \textit{originality} in one’s cares, or devotion to projects that are not simply copied from others. In this respect, "individuality" as Frankfurt understands it is like the existentialist ideal of \textit{authenticity}, without its liberty component, as I defined it earlier (see §Int.4).

Frankfurt then argues that authentic individuality or full self-determination in this sense necessarily \textit{conflicts} with liberty at a certain point. As our field of alternatives is extended, the agent may "experience an unsettling diminution in the clarity with which he comprehends who he is." Worse, if this field "has no boundaries at all," so that "every conceivable course of action" is volitionally possible, then the agent will be able to choose all of the desires, preferences, and criteria by which he makes practical choices: "It will be possible, then, for him to change those aspect of his nature that determine what choices he makes."\textsuperscript{1260} If all of the motives for choice are thus "adventious and provisional" until the agent commits himself to them, then he has no basis for deciding how to commit himself. "Under these conditions there is in him no fixed point from which a self-directed volitional process can begin."\textsuperscript{1261} Thus he is "so vacant of identifiable and stable volitional tendencies and constraints" that "the decisions he makes will be altogether arbitrary," both groundless and reversible.\textsuperscript{1262} Such a being has a kind of liberty that makes individual self-determination or authenticity impossible. The only way to avoid this is for the person to have volitionally necessary "ideals" that fix some original limits to what is volitionally possible for him.

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{1259}Ibid. p.17.
\item \textsuperscript{1260}Ibid. p.18.
\item \textsuperscript{1261}Ibid.
\item \textsuperscript{1262}Ibid. p.19.
\end{enumerate}
\end{footnotesize}
A person without such a "personal essence" would be incapable of "genuine integrity" because, again, "There is nothing that he is essentially:" he lacks the metaphysical basis from which he could be said to confer his authority on any motive, project, or course of action he takes up. Such an agent would not be a person.

This critique of the rootless agent is obviously directed against the Sartre’s nihilistic picture of the "for-itself” as a freedom that cannot be moved in choosing its original project by any motives or values unless it consents to give them motivating force, and which (as Sartre emphasizes) is therefore haunted by anguish at the arbitrariness of its groundless choice of ultimate values or highest commitments. On Sartre’s picture, the for-itself is a torturous contradiction: it desires authenticity but is condemned to a freedom that prevents this. The for-itself could acquire a stable and self-determined character without self-deception or "bad faith" only if it could by sheer fiat provide genuine grounds for its choices (as Nietzsche thought it could), but this is impossible, and so we are absurd beings.

I agree with Frankfurt that this Sartrean picture is both deeply unsatisfying and unconvincing. If it were really impossible for the choices a person makes now even to influence or incline his later choices (to say nothing of determining them), how could we actually care about what we are

---

1263See Sartre, Being and Nothingness, p.71: "This freedom which reveals itself to us in anguish can be characterized by the existence of that nothing which insinuates itself between motives and act. It is not because I am free that my act is not subject to the determination of motives; on the contrary, the structure of motives as ineffective is the condition of my freedom....this nothing is made-to-be by the human being in his relation with himself. The nothing here corresponds to the necessity for the motive to appear as motive only as a correlate of a consciousness of motives...there is never a motive in consciousness; motives are only for consciousness.”

1264Ibid: "It follows that my freedom is the unique foundation of values and that nothing, absolutely nothing, justifies me in adopting this or that particular value, or this or that scale of values....My freedom is anguished at being the foundation of values while itself without foundation” (p.76). And: "...all these commonplace everyday values, derive their meaning from an original projection of myself which stands as my choice of myself in the world. But to be exact, this projection of myself towards an original possibility, which causes the existence of values, appeals, expectations, and in general a world, appears to me only beyond the world as the meaning and abstract logical signification of my enterprises” (p.77).

1265Ibid, p.616.

1266Ibid, p.70: "But what he apprehends in anguish is the total inefficacy of the past resolution....It seemed to
going to do in the future, and thus be anguished at not being able to determine it with certainty?

Similarly, (in Frankfurt’s terms) if it were not objectively important to care about certain things, or to engage in certain sorts of caring (whatever their particular objects and aims), then how could we care about what we care about, or (in my terms) care to look for grounds for our projective self-motivations, as anguish at our groundless clearly presupposes we do? Thus Sartre’s picture is phenomenologically inadequate.

Yet interestingly, Sartre arrives at this point precisely because of a premise that he shares with Frankfurt. Like Frankfurt, Sartre assumes that absolute freedom or absolute fixity are the only alternatives at the final level of the self. Rejecting the latter, he writes:

Since freedom is a being-without-support and without-a-springboard, the project in order to be must be constantly renewed. I choose myself perpetually and can never be merely by virtue of having-been-chosen; otherwise I should fall into the pure and simple existence of the in-itself.

Reacting primarily against the behaviorists, Sartre assumes that any inertia or resistance to change that would make features of our character such as dispositions, values, desires, and other motives substantive, or let them affect or provide a basis for latter choices without being re-chosen at the time would amount to determinism. Effective motives, or factual constraints on freedom, would entirely eliminate liberty at the original level of self-determination and equate the individual with something that simply is what it is: the individual would be her innermost character in the way that

---

1266(...continued)

me that I had established a real barrier between gambling and myself, and now I suddenly perceive that my former understanding of the situation is no more than the memory of an idea, a memory of a feeling. In order for it to come to my aid, I must remake it ex nihilo and freely. The not-gambling is only one of my possibilities, as the fact of gambling is another of them, neither more nor less (p.70, my italics).

Notice here Sartre’s assumption that either there is a barrier fixed like a thing that absolutely determines the behavior, or there is no barrier at all and all choices are equally available. He leaves out the intermediate possibility of a factically inclined and constrained but still undetermined freedom which has multiple alternatives but for which none of them are equally choosable.

1267 See Frankfurt, “The importance of what we care about,” p.92: "...it is necessarily important to people what they care about....if anything is worth caring about, then it must be worth caring about what we care about."

1268 Sartre, Being and Nothingness, p.617.
Accepting this dilemma, Sartre rejects the reification of the individual and pursues the opposite horn of the dilemma to its most extreme implications. Frankfurt also accepts the dilemma, and but rejects the alternative of nihilistic liberty, and pursues his horn to its extreme implications: he is forced to equate the individual with a volitional or identificational destiny. In other words, ironically, Sartre and Frankfurt each accept the conclusion that the other found unthinkable, and reach their own conclusion just because they found the other unthinkable. Thus Sartre heaps scorn on the notion of a characterological essence that Frankfurt has adopted:

What I attempt to flee here is my very transcendence, insofar as it sustains and surpasses my [characterological] essence. I assert that I am my essence in the mode of being of the in-itself....I apprehend, or at least try to apprehend it as the original beginning of my possible, and I do not admit at all that it has in itself a beginning. I assert then that an act is free when it exactly reflects my [characterological] essence. However, this freedom, which would disturb me if it were freedom before myself, I attempt to bring back to the heart of my essence—i.e. of my self. It is a matter of envisaging the [innermost] self as a little God which inhabits me and which possesses my freedom as a metaphysical virtue....In short, it is a matter of apprehending my freedom of in my self as the freedom of another. We see the principal themes of this fiction: My self becomes the origin of its acts as the other of his, by virtue of a personality already constituted.(p.)

In other words, if my decisive identification were really fixed by absolute volitional necessities that constitute me, then I would be alienated from "me," and my deepest commitments would not in fact express my most unambiguous authority as they do. So Frankfurt’s theoretical conclusion that there must be individual volitional essences is a metaphysical fiction that fails to explain the very phenomena for which it is trying to account. But Sartre’s alternative is also implausible, as we have seen.

The only conclusion to draw from this is that Sartre and Frankfurt both set up a false dilemma. We do not need to accept that a person’s volitional character is either ultimately arbitrary or determined without liberty. These are not the only alternatives, since another model is available in which the agent never acts without an acquired character, values, and commitments that influence
the appearance of his possibilities and incline his choices, but these factual conditions that constrain
the agent still do not *determine* his choices, even to the extent of preventing him from bringing about
changes in these conditions themselves. This is the model I will develop further in the next and final
section.

VI.7. Liberty without Arbitrariness: An Existentialist Response to Four Objections

VI.7.1. Why Existentialists Must Reject ‘Linkage’ between the two aspects of choice

Frankfurt’s conceptual argument for absolute volitional necessities is simply one version
—tailored to the existential issue of decisive identification or ultimate commitments— of a more
general *argument-type* that has frequently been used against different libertarian accounts of free
agency. In the metaphysical literature on free will, it is known as the *problem of indeterminism*: if
we exercise a freedom that is incompatible with natural determinism, then how does the
indeterministic process of selecting one option out of many avoid being a *randomizing* process?
How can it be the agent’s choice, rather than something which merely happens to him? In current
literature, there are two basic approaches to resolving this problem. One is to develop a theory of
*agent-causation* which gives a causal explanation for how the decision or selection of an option can
be said to ‘flow from’ the agent.\footnote{This type of account, e.g. Chisholm’s theory of free actions, is kind of account Heidegger rejects when he
 critiques Scheler’s theory of the person as “executor” of actions. Heidegger asks what “execution” is to mean,
if not that the behavior is caused by a *person* (which makes Scheler’s account circular). Similarly, we must ask
Chisholm for an explanation of what the *agent* is that causes acts, rather than mere events. If we say that an
agent is a being that executes acts, we run into the same kind of circularity.} The other general approach is to explain choices in terms of
*intentional narratives* that rationalize how the option acted on ‘fits with’ the agent’s subjective
considerations and psychic states such as reasons, deliberation, emotions, preferences, prior
dispositions, and so on. And there are many combinations of these two approaches.

Notice that both these approaches can also be taken by someone who rejects libertarian freedom.
Plato may have believed in agent-causation while denying that we are free to act in any way other than what our intellect has independently determined to be good for us.\footnote{It is a bit more complicated for Aristotle, Augustine, Aquinas, and others in this tradition.} And Frankfurt gives an account of autonomy solely in terms of the relation of the agent’s reflexive intentional states to the intentions on which he acts. Thus these approaches do not imply libertarianism. But for libertarians, they may offer ways to preserve a sense of self-determination in libertarian choices. However, each of these approaches allows for such a combination in a way different from the other in a crucial respect. In particular, as I will argue in this section, the intentional narrative approach cannot propose to give a complete explanation of why the agent selected one option over others. On this approach, the liberty in the selection of alternatives and the sense in which the selected option ‘comes from’ the agent must be kept distinct, rather than jointly explained by a single condition.

Any promising libertarian account of freedom will acknowledge a distinction between two different types of metaphysical condition for a morally imputable action:

1. **Liberty conditions**: that some sufficiently rich range of (kinds of) alternative actions be accessible to the agent or ‘within her power’ in the circumstances.

2. **Agent-Connection conditions**: that the behavior involved in the action have the right sort of connection to the agent, or relation to her own relevant states, to count as an expression of that agent’s will, or as self-determined, or as controlled by the agent.\footnote{Different writers have used different terms for the concept which this type of condition tries to explain or interpret. Gary Watson uses "self-determination," the phrase which I have followed here. O’Connor uses "agent-control." Others, like Frankfurt, unfortunately use "autonomy" for this concept, explicitly contrasting "autonomy" in their sense with libertarian freedom. This is confusing because, since Kant, "autonomy" has been used for a freedom that combines self-determination with liberty (the term I prefer for having a rich set of alternatives). Susan Wolf, for example, uses "autonomy" in this combined Kantian sense. This proliferation and inconsistency in terminology across authors muddies the waters in the current literature, e.g. making it harder than it needed to be to compare Wolf’s and Frankfurt’s views.}

These conditions under (2) may further subdivide into two parts:

a) Those which define what it is for some motion, behavior, or event to count as an action, as opposed to a mere happening not attributable to the agent;

b) Further conditions that define what it is for an action, in addition, to have the authority of
the agent behind it, or to count as fully imputable (this allows for the possibility, for example, that some behaviors might count as intention-guided acts, while still not being fully self-determined, because the agent does not identify with the motives underlying his intentions, etc.)

As we have seen, libertarian theories need some account of the conditions under (2) in order to explain in what sense our "actions" under conditions of liberty (1) are not merely random, accidental, or arbitrary. Every libertarian account wants the alternatives among which we are "at liberty" to choose to be alternatives that are actions, not mere happenings, and some libertarian theories (those that follow Frankfurt in distinguishing actions and imputable actions) will also want them to include alternative possible authoritative or fully self-determined actions.

Different libertarian accounts will spell out these conditions in different ways. In particular, agent-causal theories at least spell out conditions 2(a) in terms of the behavior’s having some causal connection with states in the agent, or ‘the agent himself.’\(^{1272}\) By contrast, intentional-narrative libertarian accounts spell out 2(a) (and often 2(b) as well) in terms of some non-causal link between the behavior that counts as an act and a range of different types of internal states —intentions, the beliefs, desires and reasons that inform intentions, or deeper elements of "character" that pattern our motives, e.g. dispositions, higher-order volitions and identifications, commitments, cares, concerns— that lie behind the agent’s apprehension of her behavior under some intentional description, or according to one of its aspects (e.g. "entering the bar for a drink" vs. "avoiding someone approaching on the street," etc).

**Linkage:** Some libertarians believe that however the conditions under (2) are spelled out, they

\(^{1272}\)Agent-causal theories, in addition to their other failings, often seem to ignore the need for additional conditions of the 2(b) variety. I suspect that this is not simply a coincidence: for it is especially hard to imagine how, having defined action in causal terms, we could then go on to spell out a further set of conditions for an action’s being "internal" or fully self-determined in the sense that it has the full authority of the agent in it (or that the agent is fully "active" in this act) in other purely causal terms. Of course one could have an agent-causal theory about actions, and then other non-causal conditions to fill out self-determination, but given that 2(b) will almost certainly have to be cashed out in non-causal terms, it will be hard to have an integrated account of self-determination unless one also has a non-causal account of actions in terms of their being guided by (or fitting with) the relevant intentions and the reasons and motives behind these intentions.
must be linked in a certain way to the phenomenon of liberty—the phenomenon which incompatibilists hold to be incompatible with determinism. The linkage goes as follows:

There is some event, process, or act (e.g. intention-formation) $X$ by which the agent arrives at one of his multiple accessible alternatives, or "selects" one of the possible options he is at liberty to actualize over the others. Whatever $X$ is, $X$ must also function as the explanation of why this option, once selected, counts as an action, or as self-determined, or as ‘coming from’ the agent in a suitable sense.

When ‘linkage’ in this sense is satisfied, then the explanation of why the particular act that the agent does counts as an act or as self-determined in some fuller sense, will involve the same phenomena that explain why it was chosen over the other options, or (since this need not be in a justificatory sense of "why") at least how this act’s being elected out of the set of possible alternatives (the liberty-set) derived from the agent. In other words, ‘linkage’ means that what it is for an action to be selected and what it is for it to come from the agent in the appropriate senses are the very same thing. So for example, in an agent-causal account, we may be told that $X$ is a kind of causation arising directly from the agent without any other causal antecedents: it is by the agent $X$-causing a given act A that she chooses it out of her liberty-set, and it is because A is $X$-caused that A counts as an action of that agent or as agent-controlled.

Agent-causal accounts seem to be able to satisfy the linkage condition, while intentional-narrative accounts of self-determination seem unable to satisfy such a linkage requirement. In fact, the former satisfy linkage almost too well: it looks as if the agent-causal approach in general is tailor-made precisely for the sake of preserving linkage. But if this is the real motivation for taking this approach, it looks like the agent-causal agenda is driven by the need to hang onto a mistake that derives historically from the preconception deeply rooted in our philosophical heritage that libertarian choice and self-determination must at bottom be the same thing, or one single

---

1273 Though I do not have the space to discuss agent-causal theories in detail, in my opinion an existentialist has to reject such theories, since they typically begin with a notion of the person as a substance in relation to which there can be substance–event causation, rather than merely event–event causation in the world.
phenomenon, which was therefore given a single name, "voluntary action," to cover its equivocation between selection and agent-connection. The agent-causalist imagines, in his confusion, that it is the very picking between options (whose common availability to me defines my liberty) that makes the option picked my act, or that makes it self-determined, or attributable to me rather than to something else or to chance. This is the assumption that underlies the linkage requirement, and if it is a mistake, then the main reason for preferring an agent-causal account is lost.

I will be arguing in the next several subsections that this linkage requirement, though crucial to objections against libertarianism, is indeed a mistake. We can start to see why the assumption behind the linkage requirement is misguided by considering why intentional-narrative libertarian accounts of action must reject the linkage condition. All such accounts explain the conditions of self-determination in terms of the act’s being non-causally connected to, or guided by (or ‘fitting with’), various states of mind or character that provide the basis for an intentional explanation of the action, or a story tellable from the first-personal perspective about why that action was done. Since the alternatives between which we can be said to "choose" in any meaningful sense must all at least meet the conditions under 2(a), there will (at least in principle) be some available intentional account for how that action is connected with the agent for every one of the genuine options. So the available intentional explanation shows how the act is agent-connected, though it cannot provide a complete explanation for why this agent-connected possibility was selected against all the others.

But now suppose we add in linkage. In this context, linkage amounts to the demand that the fit between the behavior and the internal states which makes possible a first-person intentional account of the behavior, thus making it the agent’s action (or making it count as self-determined in even stronger senses), must also capture or constitute the selection of this act out of the liberty-set. For if action is not explained causally but in terms of a why-story, and if libertarian selection is linked to the explanation of action, then libertarian selection also cannot be explained as a causal event but
will have to amount to there being a story as to why the option selected was privileged. Yet this combination is impossible: For (by reductio),

(i) Suppose that there must be for an agent an intentional explanation (or IE) of her act A for A to count as an act of hers, or as self-determined [non-causality];

(ii) And suppose that for a story S to count as an IE of A, it must give a complete explanation of A, in other words, that it must explain why the agent did that act A as opposed to any of the other options open to her or within her power [linkage];

(iii) Then for any possible act A* that is not chosen, there is no complete explanation, and hence no IE (by ii), and hence A* is not an act, or is not self-determined (i).

(iv) It follows that only the putative action A1 that is actually chosen and enacted can be an act, or can count as agent-connected in the way necessary for self-determination (apply iii to every option in the liberty-set other than A1).

(v) And it follows from (iv) that the agent only had one possible act (or agent-connected option) open to her, and hence she was not at liberty after all.

Thus if we want to retain liberty, we must reject one of the two premises. A non-causalist about action must reject premise (ii), which is what linkage amounts to in her system. For under the requirement of linkage, the requirement of liberty (1) and the requirement of self-determination (2) collide on a non-causal account. And it is this collision, as we will see, that has convinced many critics that libertarian freedom must be an illusion.

What this shows is that libertarians who share the intuition of many compatibilists that action must be cashed out in intentional terms, since causal accounts fail to account for the actual phenomenology of action, must reject linkage, and hence reject the demand for complete intentional explanations as well. This is the only way that libertarian freedom can be preserved while also getting the account of action right. And we can reject linkage, because linkage improperly collapses two quite different sides or aspects of free action that should be held apart. Libertarian selection and self-determination may go hand-in-hand, or be married together for us, but they are still two things in a partnership, not one phenomenon. In sum:
I. To get the phenomenology right, we need intentional conditions for action, identification, and other elements involved in the self-determination or agent-control of actions;

II. The conditions of moral responsibility ultimately demand libertarian freedom (at least at the level of identification, if not at the level of first-order actions, as we have seen);

III. But the conjunction of libertarian freedom and a non-causal or primarily intentional account of self-determination is incompatible with the linkage condition;

IV. Hence we can conclude (lacking any other defense of linkage) that the linkage requirement is false.

This conclusion will be crucial in defending the existentialist approach against its host of critics in the recent literature on autonomy and moral psychology.

VI.7.2. Four Objections Against ‘Autonomy’ As Liberty

In recent work in Anglo-American moral philosophy, there has been widespread concern that an existentialist picture of the freedom required for responsibility for one’s ‘self’ (or deep motivational character) projects an impossible ideal of total self-construction, in which free choice is paradoxically reduced to complete arbitrariness. In the next subsections, I’ll review widely respected formulations of this and closely related objections by Thomas Nagel, Susan Wolf, and Charles Taylor, who have all critiqued the idea of an ‘autonomous agent.’ In the final subsections, I’ll offer an existentialist response by outlining a Kierkegaardian theory of volition and the higher-order will as simultaneously free and dispositional, which allow the existentialist to maintain liberty without arbitrariness in the shaping of our deep character or ‘real self.’

In The View From Nowhere, Thomas Nagel defines ‘autonomy’ as "the sense that we are authors of our own actions" or that we actualize one out of "some range of open possibilities... presented... on

\[\text{1274}\text{The text from this section to the end is a modified version of my paper, "An Existentialist Theory of Volition: The Paradoxes of Autonomy Resolved in a Freedom without Arbitrariness," presented at the Eastern Division meeting of the American Philosophical Association (Philadelphia, PA, December, 1997). I am grateful to Bernard Reginster for his comments and challenging questions about this paper.}\]
an occasion of action. But if autonomy involves alternate-possibilities freedom or ‘spontaneous’ indeterminacy, however, Nagel thinks it turns out to be unintelligible, because it tries to combine such liberty with acting on reasons: "A free action should not be determined by antecedent conditions and should be fully explained only intentionally, in terms of justifying reasons and purposes." But (as we have seen) these two conditions are opposed: if the agent can elect either of two mutually exclusive options, and the noncausal account of action in terms of considerations counting as reasons in "the agent’s point of view" applies whichever way she goes, then it cannot "explain" in terms of reasons why she chooses one option rather than the other:

Intentional explanation, if there is such a thing, can explain either choice in terms of the appropriate reasons, since either choice would be intelligible if it occurred. But for this very reason it cannot explain... on grounds of intelligibility, why one of two intelligible courses of action, both of which were possible, occurred. And even where it can account for this in terms of further reasons, there will be a point at which the explanation gives out. We say that someone’s character and values are revealed by the choices he makes in such circumstances, but if these are indeed independent conditions, they too must either have or lack an explanation.

Autonomy is thus incoherent because the liberty required by one part of its definition implies that the other part cannot be satisfied, i.e. that the intention-grounding considerations salient for the agent before her choice (her desires, reasons, beliefs, and so on)— can at best underdetermine her decision. And the destructive effects of liberty in the ideal of autonomy get worse, because when connected to our capacity to adopt more and more "objective" or "external" viewpoints, it leads us to "feel that in acting, we ought to be able to determine not only our choices, but the inner conditions of those choices, provided we step far enough outside ourselves." But then the ‘choice’ by which we

---

1276 Ibid, p.115.
1277 Ibid.
1278 Nagel, p.116.
1279 Nagel, p.118.
adopt our principles of choice, or determine what will count for us as reasons, desirable ends, and so
on, must be completely arbitrary. If so, then our choices to act in one way rather than another are not
only underdetermined by the available elements of intentional explanation: they are not ultimately
influenced by rational considerations, character, or values at all, since such deliberative and
dispositional elements of intentional action are themselves the product of deeper spontaneous
choices.

I have framed Nagel’s paradox of autonomy this way to bring out the fact that it combines what
are really two distinct objections, although Nagel does not explicitly distinguish them. (A) The first
is the problem of sufficient explanation: the agent’s character and rational considerations may inform
a libertarian choice but cannot by themselves provide the complete intentional explanation for the
selection of one possibility out of the available range, or the "full" intention explanation that Nagel’s
definition of autonomy requires. (B) The second is the problem of arbitrariness: if the very
authority of reasons for the agent, as well as his operative dispositions, values, and interests, all
result from still more fundamental libertarian choices, then both the agent’s self and the actions
which flow from his character and deliberative engagement will be completely arbitrary, like blind
chance. Notice that this second problem also divides into two parts: (B1) the basis for the objective
authority of action-guiding principles and values, and (B2) the origin of motivating reasons, desires,
dispositions, and commitments that make up our volitional ‘character.’ In addition to these
difficulties, Nagel also describes a third and more familiar problem (C): "From an external
perspective...the agent and everything about him seems to be swallowed up by the circumstances of
action." 1280 While (A), (B1), and (B2) are difficulties internal to the very concept of autonomy, or
difficulties in making sense of "what intelligible belief is undermined by the external view" of our
nature as sociobiological beings, (C) is just the problem that autonomy (even if it can be made

1280Nagel, p.114. See also p.123.
internally intelligible) is undermined by this external view.\textsuperscript{1281}

These and other related difficulties with autonomous accounts of human character and agency have been emphasized in other well-known works following Nagel’s contribution. In a landmark study, Susan Wolf has analyzed ”The Dilemma of Autonomy,” in a fashion reminiscent of Nagel, as a problem of correcting inherently inconsistent intuitions about ”What kind of beings” we must be ”if we are ever to be responsible for the results of our wills.”\textsuperscript{1282} Being responsible requires that we have "a will" or a capacity to form intentions that can effectively guide our actions,\textsuperscript{1283} and that this be a "suitably intelligent will," i.e. "a will whose content can be informed and governed by the relevant considerations" in the circumstances. Thus if "unavoidable ignorance or inexperience" prevents us from recognizing the "relevant sorts of features” that weigh for or against different options in the situation, then we are exempt from responsibility.\textsuperscript{1284} These two conditions are similar to the intentional explanation component of Nagel’s concept of ”autonomy,” and are central to the ”Reason View” which Wolf herself defends later in the book.\textsuperscript{1285} However, these conditions still appear to be jointly insufficient for responsibility, because in cases such as duress and post-hypnotic suggestion, agents’s wills "are effective and relevantly informed,” but "the content of their wills results from something else, and this something else is such that the agents are powerless to choose or resist it."

\textsuperscript{1281}Nagel, p.117.

\textsuperscript{1282}Susan Wolf, Freedom Within Reason (Oxford University Press, 1990), ch.1, ”The Dilemma of Autonomy,” p.4.

\textsuperscript{1283}Wolf, p.7: "It is obvious enough that one condition of responsibility involves the possession of a will. In other words, only an agent who has a will—that is, desires, goals, or purposes and the ability to control her behavior in accordance with them—can be responsible for anything at all." Notice that will in this sense is roughly equivalent to Frankfurt’s notion of a first-order will that guides bodily acts. Wolf adds that a necessary (although not sufficient) condition of responsibility for ”events and properties” is that ”her will is or could have been effective in determining the existence of these events or properties” (p.7). This is similar to Frankfurt’s condition that we are not responsible for behavior that is not guided by our intentions —for if the intention is blocked from being effective, the behavior is not an ”act” at all.

\textsuperscript{1284}Wolf, p.8.

\textsuperscript{1285}See Wolf, ch. 4.
Thus, as Harry Frankfurt has also stressed, we can have "desires that move us but with which we do not identify." It is this concern, as Wolf sees, which motivates the problematic libertarian component in autonomy, or:

...the requirement that the agent’s control be ultimate—her will must be determined by her self, and her self must not in turn be determined by anything external to itself. This last condition I shall call, after Kant, the requirement of autonomy.

I call this a libertarian condition because Wolf builds into it the Kantian idea that we can identify with the desires or motives on which we can be said to "act intentionally" only if we can choose either to identify or not to identify with them. In other words, if we think of our "self" in Frankfurt’s terms as constituted by our "second-order volitions" or volitional "identifications," then autonomy (now construed as liberty of the higher-order will) ensures that we are the sole authors of our self: our identifications cannot be externally determined, and they must control which desires and considerations move us in forming intentions to act, or become our first-order will, if we are to be responsible for our actions.

So conceived, the requirement of autonomy encounters two problems, according to Wolf. First,

---

1286 Wolf, p.10.
1288 Wolf, p.11.
1289 Wolf, p.10.
1290 We should note, however, that Wolf’s terminology here follows a potentially misleading trend in the recent literature. Like Nagel, Wolf uses "autonomy" for what Kant would have called spontaneity, since it does not refer to motivation by pure practical reason or by the formal universalizability of the maxim or intention on which we act. For Kant, of course, there is spontaneity in "autonomous" action in his sense, but human "heteronomous" action is also spontaneous and morally imputable. On this point, see Henry Allison, "Spontaneity and Autonomy in Kant’s Conception of the Self," in The Modern Subject: Conceptions of the Self in Classical German Philosophy, ed. Karl Ameriks and Dieter Sturma (SUNY Press, 1995): 11-30. The original version of this paper was presented at the Conference on German Idealism, University of Notre Dame, April 8, 1994), and the essay is reprinted in Allison, Idealism and Freedom.
1291 See Frankfurt, "The Freedom of the Will and the Concept of a Person," in The Importance of What We Care About, p.18.
careful observation reveals that we hardly ever simply create our own desires; although the manipulation is not normally as conspicuous as in cases of duress, an external view suggests "a picture of ourselves as creatures whose desires are a result of some combination of our heredity and environment, who try to satisfy our desires as well as we can be acting as our situations demand." This is problem (C), the external view or sociobiological objection, as we might call it: it is not a problem with the coherence of autonomy in principle as a condition for responsible action, but rather the fear that this condition is too demanding, since it cannot be met by situated creatures like us.

The second problem is internal to the very idea of autonomy. Autonomy is inconsistent with the agent’s motivations being controlled either by "external forces" or by "nothing at all." But then, if the fact that her will is "up to the agent herself" means that there should be a complete intentional explanation for why she performs any of her actions, then autonomy will be impossible:

For it seems that about any agent and any act whatsoever we can ask for an explanation of why that agent performs that act. And though we may begin to answer this question in terms of features internal to the agent, we can always press beyond these beginnings and ask why the agent possesses these features....Eventually we will reach a set of features that must be explained by facts external to the agent [i.e. unchose] or our explanation will simply come to an end, with the understanding that the agent’s possessing these features is either a random occurrence or a brute, inexplicable fact.

This dilemma combines Nagel’s sufficient reason and arbitrariness objections (A and B). Intentional explanation is incomplete (in a different respect than Nagel indicated), because we must explain why the agent has the personality (commitments, emotions, and dispositions), and the desires, reasons, and values from which intentional explanation begins. If these features themselves are externally caused, then according to autonomy, we could not be responsible for our actions, so they must have an internal explanation. But what factors are left that would give these features a purely internal

\[1292\] Wolf, p.12.
\[1293\] Wolf, p.13.
\[1294\] Wolf, p.13.

508
origin? If we say that they are primitive or chance events, they are internal but not explained. If we say that they originate from choices, then these choices either have intentional explanations, or they do not. If they do not, then they are arbitrary (like chance); but if they can be explained in terms of further reasons, dispositions, values, and so on, then the dilemma will apply again to these features.

In order for an agent to be autonomous, it seems, not only must the agent’s behavior be governable by her self, her self must in turn be governable by her self—her deeper self, if you like, and this in turn must be governable...ad infinitum. If there are forces behind the agent, so to speak, making the agent what she is, then her control of her behavior is only intermediate, and therefore superficial. But if there are no forces behind the agent making the agent what she is, then her identity seems to be arbitrary.¹²⁹⁵

To grasp what is going on with the notion of ‘deeper selves’ in this dilemma version of the arbitrariness objection, however, we have to return to the Frankfurtian idea mentioned earlier that persons are beings capable of identifying through second-order volitions with the desires or motives on which they act (the intentions which guide their bodily behavior). As one possible response to this dilemma, Wolf considers the "Real Self View," which develops the Humean intuition that the freedom which is required for moral responsibility is "agents’ normal abilities to control their behavior in accordance with the values and choices of their deepest selves."¹²⁹⁶ One account of such a Real Self would equate it with higher-order volitions, but because of apparent problems with this explication,¹²⁹⁷ Wolf prefers Gary Watson’s analysis that an agent’s real self is constituted by her

¹²⁹⁵Wolf, p.14. She sums this up by saying that "The idea of an autonomous agent appears to be the idea of a prime mover unmoved whose self can endlessly account for itself and for the behavior that it intentionally exhibits or allows. But this idea seems incoherent, or, at any rate, logically impossible (p.14).

¹²⁹⁶Wolf, p.28.

¹²⁹⁷As Wolf puts this option, since the agent whose desires do not follow from her Real Self is "alienated from her will," we might try to say that the responsible agent "acts according to her will, and ...wills according to her choice" [where "choice" = second-order volition] (p.29). Against this construction, she objects (1) that "we do not naturally think of our ordinary actions as resulting from a series of choices," and (2) that an agent who is alienated from her first-order choice may be alienated from her higher-order choices as well” (p.30). Though I believe Wolf’s critique of Frankfurt’s involves several confusions (e.g. conflating identification with our desires with causing them to be the desires on which we act) and buys too quickly into Gary Watson’s argument that it entails a regress in orders of volitions, the question of how to reconstruct Frankfurt’s hierarchical model to avoid these problems is not my theme here and now, so I will leave this debate aside.
"values," or judgment-informed desires (as opposed to mere preferences or motivationally operative desires). In particular, the actions attributable to the agent’s real self are those arising "out of her valuational system" as a whole.

But whether it is explained this way or in terms of Frankfurt’s higher-order volitions, the Real Self View rejects the liberty to determine our ultimate character and values, the liberty which led to the arbitrariness problem: instead of ‘autonomy’ in Wolf’s sense, responsibility on this view only requires that the agent’s "actions are governed by her will and her will is governed by her unalienated, real self," the "self with which she is properly identified." Thus as Wolf says,

..this view does not require the agent to be endlessly accountable to herself —that is, does not require that her self be governable by her self ad infinitum. It requires that an agent have a real self, and that she be able to govern her behavior in accordance with it. But it does not matter where her real self comes from, whether it comes from somewhere else or nowhere at all. Let us call this the Real Self View.

But precisely for this reason, as Wolf recognizes, the Real Self View remains problematic: "For we sometimes have reason to question an agent’s responsibility for her real self," if she seems through no fault of her own to be controlled by a corrupt valuational system. "Deep responsibility" (both in the ‘narrow’ moral sense and in terms of virtues and vices) seem to be lacking when real selves or entire valuational systems are distorted by "comprehensive insanity, psychological conditioning, and

1298 Wolf, p.31.
1299 Wolf, p.32.
1300 Wolf, p.30.
1301 Wolf, p.35.
1302 As Wolf explains, "deep responsibility" is a judgment of the "moral quality of the individual herself in some more focused, noninstrumental, and seemingly more serious way" than simply attributing causal roles to her (p.41). Moreover, "While moral responsibility is probably the least controversial and best-examined species of deep responsibility, there may be other species of deep responsibility," such as the attributions of "qualities like courage, patience, arrogance, closed-mindedness, which are not moral qualities in the narrow sense of that term that connects morality to benevolence and impartiality" (p.41).
dramatically deprived childhoods," for example.\textsuperscript{1303} The problem with the Real Self View, then, is that in trying to provide non-libertarian conditions for our responsibility for actions in terms of our valuational or volitional character, it fails to consider deeply enough the question of how we can be responsible for this inner character itself. Suppose, however, that we try to solve this problem by bringing autonomy back again at the level of the real self (or, in Frankfurtian terms, introduce liberty into the higher-order will). This strategy, which I believe is the closest to describing the considered existentialist position, Wolf calls the "Autonomy View." The proponent of this view has on his side that values (or higher-order volitions) "differ from other desires in being motivations that we not only have but care about having" and view as good; while we may just passively find other desires in ourselves, our values seem to involve our positive "choice" at least in the counterfactual sense that we "would embrace" them if the question self-consciously arose.\textsuperscript{1304} The distinction here is similar to Charles Taylor’s distinction between "strong evaluation," in which I identify with being "a certain kind of person," and "simple weighing" of desires and preferences of different strengths.\textsuperscript{1305} As Taylor says, the very term "evaluation" connotes this active mode

...our [strong] evaluations emerge from our activity of evaluation...This active sense is conveyed by Frankfurt’s formulation where he speaks of persons as exhibiting ‘reflective self-evaluation’ that is manifested in the formation of

\begin{flushleft}
\textsuperscript{1303}Wolf, p.44.
\textsuperscript{1304}Wolf, p.49.
\textsuperscript{1305}See Charles Taylor, "Responsibility for Self," from The Identities of Persons, ed. A.O. Rorty (University of California Press, 1976), reprinted in Free Will, ed. Gary Watson (Oxford University Press, 1982): 111-126, p.113. As Taylor puts it, "Someone who evaluates non-qualitatively, that is, makes decisions like that of eating now or later, taking a holiday in the north or south, might be called a simple weigher of alternatives. And the other, who deploys a language of evaluative contrasts ranging over desires we might call a strong evaluator" (p.116). Wolf’s notion of "valuing" is similar to Taylor’s "strong evaluation." As she says, "The picture that develops out of attention to these experiences is one according to which our valuing selves may seem to rise above our more brutish, desiring selves. As valuers, it seems, we are able to take stock and survey our desires and dispositions, rejecting some perhaps as utterly worthless, and ranking others in ways that may take account of but need not ultimately reflect their relative motivational strengths" (Wolf, p.50). Thus the ranking in valuation is like Taylor’s contrastive or qualitative evaluation of options, as opposed to the competition of desires by their quantitative strengths in the ‘economy’ of simple weighing.
\end{flushleft}
second-order desires.\textsuperscript{1306}

Yet, Wolf insists that if we take this route, our real or valuing selves will not be autonomous selves, because though our ‘values’ (or higher-order volitions) stand ‘above’ our brute or unjudged motivational preferences, they cannot be determined by libertarian choices:

[She says]...what path we do take is hardly accidental, for our choice arises out of, and is explained by, our reflections and deliberations. [Thus] It is doubtful... that the sense in which we seem to have a choice about our values is properly characterized as an experience of autonomous choice. For the idea that we choose what to care about and what to do in a way that is independent of, or at any rate not necessarily determined by, the force of our unvalued passions may yet be interpreted in a way that suggests that our choices are determined by something else...The sense in which we cannot help having the desires we have but can help what values we have may then be interpreted as a reference to the fact that our values, unlike our mere desires, can be controlled or chosen in accordance with Reason.\textsuperscript{1307}

Wolf argues that since the deliverances of reason are not controlled by choice,\textsuperscript{1308} if the "valuing" or ‘real self’ is thus identified with the rational self—or in Frankfurt’s terms, if our higher-order volitions are those formed solely by our faculty of Reason—then our inner self is not autonomously formed. Autonomy requires liberty or alternate possibilities with respect to reason: "The autonomous agent must be one who is able to act in accordance with Reason or not."\textsuperscript{1309} And from such liberty, Wolf follows Nagel in inferring arbitrariness or indifference:

Since a radical choice must be made on no basis and involves the exercise of no faculty, there can be no explanation of why or how the agent chooses to make the radical choices she does.\textsuperscript{1310}

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{1306}Taylor, p.118.
\item \textsuperscript{1307}Wolf, p.51.
\item \textsuperscript{1308}Wolf, p.52. Yet see Frankfurt’s rather different analysis in "Identification and Wholeheartedness," reprinted in The Importance of What We Care About: 159-176, esp. p.175-6, where Frankfurt argues that practical reasoning itself inevitably involves making decisions which are not determined by the available considerations.
\item \textsuperscript{1309}Wolf, p.54.
\item \textsuperscript{1310}Wolf, p.54.
\end{enumerate}
\end{footnotesize}
As a result, like Dennett, Wolf thinks that autonomy is not a freedom worth wanting. To want it, she says, we would have to "want the ability to act irrationally," and be able to "view the possibility of acting irrationally as potentially desirable." But desiring or approving of this ability could not be justified or motivated by any reason. Since this is in some sense a distinct objection, let me label it (D): even if autonomy is a coherent possibility, it would be unintelligible to want it, because it would be worse for the agent to have such autonomy.

In his own searching analysis of these questions, Charles Taylor agrees with Wolf that (i) the problem with which we are concerned not simply responsibility for one’s outward actions, but responsibility for our inner character, and (ii) that it cannot be explained by radical choice:

...responsibility for oneself [...] goes along with this notion of the agent as a strong evaluator. Naturally we think of the agent as responsible, in part, for what he does; and since he is an evaluator, we think of him as responsible in part for the degree to which he acts in line with his evaluations. But we are also inclined to think of him as responsible in some sense for these evaluations themselves.

Yet this responsibility for our strong evaluations, or (in Frankfurt’s terms) for the inner character constituted by our pattern of higher-order volitions, cannot be grounded in an absolutely arbitrary

---

1311 Wolf, p.55. Compare Daniel Dennett, *Elbow Room* (MIT Press, 1984), p.134, where he argues that the flexibility we want is not the libertarian ability to do otherwise in the very same situation, but only to do otherwise when the situation is different in relevant respects (respects that reason could take account of). Instead, what we want is the possibility of "a future in which our deliberation is effective" in guiding our actions (p.139).

1312 Wolf, p.56.

1313 Of course (D) is really an extension of (B), the arbitrariness objection. Yet like (C), it is an objection that is *external* to autonomy, and compatible with assuming that autonomy is a coherent possibility in itself. Whereas (C) says that autonomy would then be incompatible with the sociobiological context of the creatures to whom we would attribute it, (D) says that autonomy would then be *impractical* or that it would be self-destructive to desire it. While (C) derives its force from our intuition that the freedom involved in moral responsibility must be exercised in the realm of human facticity, with all the influences of biology, culture, and circumstantial pressures, (D) derives its force from our intuition that the freedom involved in moral responsibility must be a large part of what gives us *dignity* as persons, or makes us of qualitively higher (or even incomparable) value, rather than making us worse off. There is the potential for a kind of sophistry here, however. For even Augustine, who would no doubt have been sympathetic to Wolf’s and Taylor’s reactions to Sartre and Nietzsche, thought that we would have been even more “valuable” (in his neoPlatonic ontological hierarchy) if we were confirmed angels, who could not sin, but that our freedom as an “intermediate good” still made us incomparably more valuable than lower animals.

1314 Taylor, p.118.
choice, as Sartre (following Nietzsche) urges us to believe:

But to say that they [our strong evaluations] ultimately repose on our espousing them is to say that they issue from a radical choice, that is, a choice which is not grounded in any reasons. For to the extent that a choice is grounded in reasons, these are simply taken as valid and are not themselves chosen.\textsuperscript{1315}

\section*{VI.7.3 Outline of an Existentialist Defense Against these Objections}

\textbf{B1.} In a previous article,\textsuperscript{1316} I have begun to address these objections by showing that Kierkegaard sought a way out of this very dilemma. On his account, we have \textit{liberty} in determining our higher-order will or shaping our ‘real self,’ but this does not entail arbitrariness, or the ‘Sartrean’ freedom\textsuperscript{1317} with which existentialism is so often identified. SK’s famous ‘choice to become a chooser’ refers to the decision by which we start to form higher-order volitions, thus giving \textit{personal importance} to ethical distinctions of whose objective significance we are already have some \textit{cognitive} awareness. Thus this ‘choice’ does not ground the \textit{authority} of the value-contrasts that inform ‘strong evaluations’ in Taylor’s sense, but rather \textit{appropriates them} or gives them first-personal significance. This is true even if this basic choice to engage in volitional identification may then affect our cognitive access to ethical principles and moral distinctions, e.g. either by strengthening or weakening our awareness of their authority.\textsuperscript{1318}

This provides the basis for a response to objection (B1): considerations that are reasons for action from the agent’s perspective do not \textit{become} objective reasons through any radical choice made for no

\textsuperscript{1315}Taylor, p.118.


\textsuperscript{1317}Meaning the notion of arbitrary or indifferent liberty which these critics of autonomy attribute to Sartre—however fair or unfair this may be in the context of his writings as a whole and the historical development of his views during his lifetime.

\textsuperscript{1318}For Kierkegaard’s view (as becomes especially clear in \textit{The Sickness Unto Death}) is that how we respond to our initially abstract and inadequate understanding of good and evil affects the development of our conscience, so that it is even possible to be responsible for having deadened one’s moral sense. The cognitive and volitional sides of our relation to the ethical are reciprocal in affecting each other.
reason at all; rather, the agent *identifies* with standards, principles, and evaluative distinctions whose normative status is (at least provisionally or abstractly) accessible to him independently of the choice by which he incorporates them into his real self.\(^{1319}\)

Yet the critic of autonomy will surely say that if we have to identify with reasons or value judgments to give them more than the partial or abstract cognitive significance they already possess for us, then the identification itself will be entirely unmotivated, or arbitrary. However, this does not follow. First, the Kierkegaardian existentialist can argue in response that our understanding of the reasons and evaluations *itself* motivates their personal appropriation or incorporation into the agent’s ‘real self.’ Given cognitive recognition of these considerations, it is simply false to say that the identification has no *rational* basis. Of course the decision *not* to identify with ethical ideals and distinctions is also open to the agent, and could only have a basis in a kind of irrationality (which Kierkegaard analyzes in terms of ‘sin’). But the fact that neither of these bases already present in the agent *determines* her identification does not mean that the agent’s choice in her higher-order will is ‘from nothing’ or arbitrarily suspended between its alternatives: rather, it will be *more inclined* to one or the other, depending on a host of factors (e.g. how explicit or developed her cognitive recognition of rational and ethical considerations is, how satisfied she currently is with living in the stream of immediate emotions without action by strong evaluation, etc.).

(A). At this point, the force of objection (B1) is reduced to that of objection (A): namely, that reasons which would figure in an intentional explanation of the agent’s action cannot *sufficiently* explain or *determine* her choice to go one way rather than another. But now objection (A) is open to the Kantian response that ‘freedom’ is a *practical faktum* that cannot be made intelligible from the external perspective of theoretical reason. In other words, (A) is a pseudoproblem that comes from an overly theoretical conception of intentional explanation. If we demand that the intentional

---

\(^{1319}\)Though the decision to identify with Reason in this sense may motivate a more involved deliberation than would otherwise occur in our speculative reasoning alone, prior to personal appropriation.
account of an action form an ‘explanation’ in the scientific sense, making the choice completely intelligible from a third-party perspective, then of course there will be no room for liberty: it must either be complete or ultimately arbitrary.

Thus the false dilemma that exercises both Frankfurt and Sartre (as we saw in §6.4) stems from what I called the requirement of ‘linkage’ between libertarian selection and self-determination in the explanation of action (§7.1 above). Linkage implies that the intentional explanation of a choice must be complete to be an explanation at all, or to count as connecting the act to the agent. But the linkage requirement is false, and complete intentional explanation is the wrong demand, since both identification and the ‘real self’ are essentially first-personal phenomena: as Kierkegaard argues in The Concept of Anxiety, the qualitative "leap" (or commitment of the agent in identification) is neither necessitated by the inner states which make it possible, nor is it "an act of abstract liberum arbitrium" or radical indifference. ¹³²⁰ The lack of a "full" or complete psychological explanation for the final projection of the agent is not a problem, since only states which condition without determining the choice can exist as objects for our objective understanding.¹³²¹ Nor is this just to shroud the whole matter in dubious mystery. Viewed from the personal standpoint of identification, the fact that our rational considerations and other desires underdetermine our choice of character really does not make them totally arbitrary to us. As Wolf herself notes,

¹³²⁰Note the Kant thought that liberty as such pure spontaneity was the only sense in which practical freedom could be grasped from the theoretical standpoint. External reason thus only grasps the abstract form of liberty, not its inner dispositional context nor the intentional content in which its leaps occur (i.e. as ‘maxims’).

¹³²¹Søren Kierkegaard, The Concept of Anxiety, tr. Reidar Thomte (Princeton University Press, 1980), p.49. As SK’s polemical pseudonym puts it, "To want to give a logical explanation of the coming of sin into the world is a stupidity that can only occur to people who are comically worried about finding an explanation" (p.50).
Thus objection (A) collapses into (C): (A) turns out to be a *symptom* of the external perspective that requires a complete and objective (or third-personal) explanation of action — the perspective from which we seem to be sociobiologically determined results of ‘heredity and environment.’ It will be useful to approach (C), however, by first seeing what is wrong with Wolf’s last objection (D): namely, that desiring autonomy would be tantamount to valuing action contrary to reason.

**D** First, this argument relies on the assumption that we could have "the ability to act in accordance with Reason" without *thereby* also having the ability to act against Reason, i.e. that these ‘two’ abilities are not just nominally distinct but ontologically separate, that one is not the ‘price’ of the other. The existentialist can contest this assumption, and argue (with some classical support) that in order for us to *act* on the deliverances of practical reasoning, even when they are unequivocal these deliverances cannot simply *cause* us to act on them. Second, and more crucially, even if the abilities to act from and against reason are separable, valuing the *ability* to act against reason does not entail valuing irrational action itself. As Wolf herself notes, one might value this ability, but not its exercise, because one might value the *contingency* (rather than necessity) of one’s acting on reason when one follows its dictates.  

Wolf does not acknowledge the importance of this modal distinction when she abruptly dismisses this point with the remark, "Why should one want an ability that one never wants to exercise?"  

We can want the ability precisely because our *not exercising it* 

---

1322 Wolf, p.48. Indeed, I would suggest that even when we only seem to have one feasible option, or one that makes any sense to choose, our awareness (sometimes only subconscious) that we could still rebel against making it makes the choice *subjectively underdetermined* by the factors that, from an external point of view, seem objectively to determine it.

1323 Wolf, p.57: "For though she might have every intention of saving the child, she might not like the idea of *having* to save it."

1324 Wolf, p.57.
then becomes contingent and, as a result, potentially virtuous.\textsuperscript{1325}

Wolf tries to forestall this line of response by objecting that it implies a "conception of Reason as alien to oneself," which her specifically normative conception of Reason rules out.\textsuperscript{1326} But the function of reason in identification or the constitution of the self is a phenomenological question, which cannot be resolved by stipulation. And like Eleonore Stump, who argues for reconceiving higher-order volitions as rational evaluations of the good,\textsuperscript{1327} Wolf’s inclination to equate the ‘real self’ with the evaluative dispositions that express our Reason —i.e. our judgments of the true and good— is in fact phenomenologically motivated by the familiar regress problem against Frankfurt’s original hierarchical account of identification. To avoid this problem, as Stump explains, we must account for why "an agent cannot be a passive bystander to his second-order volitions," or why he is identified with them without needing them to be approved by yet-higher order desires.\textsuperscript{1328} Linking higher-order volitions to practical reason, as in Taylor’s account of strong evaluation, seems to accomplish this.

However, as Frankfurt has consistently argued, evaluative approval and disapproval do not by

\textsuperscript{1325}I am assuming here that action cannot be virtuous if it follows automatically from features of the agent’s makeup whose ‘identification’ with her she did not bring about. I believe this is true to Aristotle’s conception of virtues as proper dispositions of choice, rather than mere habits of outward action (which may nevertheless be involved in bringing us to make the choices and acquire the motives that figure essentially in virtuous action). On this conception, even if it is more virtuous to act on the right motives without hesitation or reserve, or to have a very strong disposition to make the right choices (which makes opposed choices difficult for the virtuous agent), virtue requires that action on nonvirtuous motives not be impossible for the agent. However, I realize that this conception, which fits with my view of existentialist freedom, is not everyone’s conception of virtue. Wolf clearly disagrees with it, since she describes the agent whose virtuous action is not necessitated by her reason as by definition less virtuous, and describes virtue in the response to an emergency as requiring a "near-reflex action." These descriptions, however, are highly debatable, and the Kierkegaardian existentialist will have a coherent alternative picture of virtue to offer.

\textsuperscript{1326}Wolf, p.58.

\textsuperscript{1327}See Eleonore Stump, "Sanctification, Hardening of the Heart, and Frankfurt’s Concept of Free Will," in Perspectives on Moral Responsibility, ed. John Fischer and Mark Ravizza (Cornell University Press, 1993): 211-234, p.216. As Stump revises Frankfurt’s account, the ‘real self’ is solely identified with the agent’s intellect: "an agent forms second-order desires by reasoning (rationally or otherwise, consciously or not) about his first-order desires; and a second-order desire is a direct result of an agent’s intellect representing a certain first-order desire as the good to be pursued" (p.219).

\textsuperscript{1328}Stump, p.219.
themselves constitute identification or alienation. \(^{1329}\) For we can recognize the *normative* status or objective authority of a ‘strong’ value-judgment, and yet regard this evaluation neutrally, as if from a speculative distance, giving it only third-personal significance. The existentialist will maintain that giving the evaluation *first-personal* motivating significance, which changes it from a cognition into a volition, requires a further act of deciding to identify with the evaluation or the standards from which it proceeds, i.e. actively involving the evaluative ordering they project in one’s inner commitments. As Frankfurt notes, it is only such decisions of higher-order volition, "unlike desires or attitudes," which "do not seem to be susceptible both to internality and to externality" in relation to our real selves. Thus, although practical deliberation and rational evaluation of different possible motives or even types of *character* may be necessary components of higher-order volitions or real-self commitments, they cannot by themselves *constitute* our identifications. Thus even if reason in the normative sense *determines* our strong evaluations, it does not thereby also determine our *self* or our identification with being a given kind of person.

\(\text{(C)}\) This crucial revision in our understanding of ‘identification’ or the real self in turn suggests the possibility of a response to objection (C) which I believe Wolf overlooks. In response to the problem that the liberty in autonomy seems incompatible with our sociobiological reality, the existentialist can concede that our freedom is indeed conditioned by many kinds of *facticity* —from our inherited cultural values, our social systems and historical situations, to biologically given impulses and emotional reactions to contingencies of our environment— that constrain our available options in action and influence our character. Yet, although we may not generate the different and often conflicting desires we experience in the varying circumstances of life, our "higher-order" character is not determined by our facticity, but by the volitions through which we *identify* with some of our factically conditioned options rather than others —volitions which are not determined by our

\[^{1329}\text{See Frankfurt, "Identification and Externality," reprinted in The Importance of What We Care About: 58-68, esp. p.65.}\]
facticity. In Frankfurt’s example, although coercion from neighborhood bullies and our biological predisposition may conspire to produce our additive desire for cocaine, the existentialist maintains that we remain at liberty to be willing or unwilling addicts in our higher-order will.

Wolf does not entertain this reply to the sociobiological objection because she conceives our ‘real self’ as simply constituted by a "system of values" or rational evaluative dispositions, and, as we saw, in some cases it seems that even these strong evaluations may be determined almost completely by one’s heredity or the environment. On the Frankfurtian interpretation, however, a determinism of evaluations is not sufficient to determine identifications, since these two are distinct. The existentialist recognizes in identification a potential to resist evaluations that are forced on us, whether from psychoses, or from depravation, abuse, or conditioning. This does not deny that people in different circumstances may face unequal difficulties in autonomously determining their ‘real selves,’ or in identifying with morally good motives for action, but it does eliminate the possibility of inward selves for which the person cannot be held accountable at all. For if the agent is externally determined to adhere to any set of value-judgments, then this adherence cannot constitute a volitional identification at all. When liberty is introduced into the determination of the real self, the possibility of ‘real selves’ for which we cannot be held accountable is avoided.

VI.7.4 Conclusion: The Middle-Soul Concept of Will

Behind these (admittedly partial) answers to the foregoing objections stands the older ‘intermediate’ theory of volition (see Chap. I §3), which denies the separation of its cognitive and libertarian aspects that Nagel and Wolf assume in formulating their dilemmas. Volition involves practical reason, but is not exhausted or determined by it. On the existentialist hypothesis, the extra motivational element in volition is supplied by projective motivation. This self-initiated motivation, which is generally also a libertarian choice between nonhypothetical alternate possibilities,
completes a ‘volition’ and distinguishes it from the reasoning it involves or the grounds upon which it was projected: the volition is essentially underdetermined by the cognitive considerations through which it is nevertheless distinguished from mere affects. Precisely from this intertwining of reason and liberty, ‘volition’ gets its distinctive role as a power of the middle part of the soul in the classical sense —standing between objective, third-personal reason and first-personally motivating desire. Volition, as distinct from both pure reason and noncognitive or ‘emotive’ desire, is constituted by a rationality that becomes practical through its nonhypothetical modal openness to alternate possibilities.

In this structure, however, the ‘modal reach’ in volition can never be a liberty of indifference: the reasons, concerns, dispositions, etc. informing the volition affect both the range of options which are volitionally possible and the ease or difficulty with which different possibilities are available within this range. In other words, the different options within the range appear more or less possible for the agent, while many logically possible choices are completely excluded from the sphere of options volitionally possible for the agent. Thus the ‘liberty’ in volition never confronts a bare series of logical possibilities laid out on an equal par before it: its modal range is thoroughly shaped and colored by the considerations and character-features that figure in an intentional explanation of action. On this model, it is clear that when a volition is formed by closing on one possibility, it can never be described as arbitrary, or as a ‘radical choice’ uninformed by any cognitive criteria or preexisting character in the agent. Yet the rational deliberations and evaluations involved in volition cannot by themselves bring about the purely cognitive closure which contemporary opponents of autonomy seem to hope for.

(B2) This last point brings us back to the dilemma in objection (B2), which held that any ‘character,’ ground projects, or deep commitments that could motivate our choices must themselves either be passively given to us or elected in an utterly groundless choice. The Kierkegaardian
existentialist views this as a false dichotomy, since, as I argued in my earlier article, we never start out without any inner character or higher-order will at all, and yet the existing state of our ‘real self’ never determines the choices of higher-order will. Moreover, these choices then alter that inner character itself, forming a new inward self which again affects but underdetermines the next set of deep decisions about what kind of person to become or what kind of life to lead. The effect of these decisions is to form dispositions at the level of the higher-order will itself. As Kierkegaard says, it is this unity of disposition and freedom that distinguishes a "repetition" in which there is ongoing "originality" —a continued inward choice to sustain our commitment— from the outwardness of a mere "habit." Such ‘dispositions of repetition,’ like the rational judgments in strong evaluation, will limit the range and condition the relative availability of our future volitional possibilities, but still without determining the ‘leap’ by which we actualize them.

The solution envisioned here is similar to Charles Taylor’s notion of a "radical reevaluation" of our most basic strong evaluations, which is nevertheless situated in the inner character we already have, but undertakes an "articulation" of our "deepest unstructured sense of what is important" that is not completely determined by the prior facticity of our existing commitments and volitional dispositions. Thus without stepping entirely outside ourselves, we can still autonomously shape the deepest level of our character.

In sum, the identifications by which we form our inner self, although not determined by the rational deliberation they involve, are never arbitrary, either in the sense of being unconnected to objective criteria or in the sense of being uninfluenced by one’s existing character. Thus liberty of

---

1330 On this point, see §§ V & VII of my article on "The Meaning of Kierkegaard’s Choice Between the Aesthetic and the Ethical" (op. cit), in which I explain the aestheticism is the default position for SK, but nevertheless that the aesthete can be brought to the point where she must make a choice that will alter the aesthetic identification of her highest-order will.

1331 See Kierkegaard analysis of ‘earnestness’ in The Concept of Anxiety, p.149.

the higher-order will *does not entail indifference or ‘radical choice’* in Taylor’s, Wolf’s, or Nagel’s sense. Existential freedom is thus not ‘autonomy’ in the abstract sense that its critics attack. These suggestions, while perhaps not sufficient to dismiss the foregoing objections (A)-(D) entirely, at least show that the philosophical resources of existentialism to answer them are far greater than has been acknowledged in recent debates.
Bibliography

This bibliography includes works cited in the dissertation and others used as background for its themes and discussions. The references are arranged alphabetically by author; multiple books and articles by the same author are listed in order of the date of the latest translation, or the first date of publication in the original language (if no translation is cited). Edited and co-authored works are listed at the end of entries for each author. In some cases, references to articles include both original publication information followed by a volume in which the article is reprinted: in such cases, citations in footnotes will usually refer to the page numbers of the reprinted article in the collection, unless stated otherwise.

Allison, Henry. "Spontaneity and Autonomy in Kant’s Conception of the Self," in The Modern Subject, ed. Ameriks and Sturma (below): 11-30. The original version of this paper was presented at the Conference on German Idealism, University of Notre Dame, April 8, 1994), and the essay is reprinted in Allison, Idealism and Freedom.


Ameriks, Karl. Kant’s Concept of Mind (Oxford University Press, 1982).


Ameriks, Karl and Dieter Sturma, eds. The Modern Subject: Conceptions of the Self in Classical German Philosophy (SUNY Press, 1995).


Anderson, Thomas C. *Sartre’s Two Ethics: From Authenticity to Integral Humanity* (Open Court, 1993).


Aquinas, Thomas. *Summa Theologica* will be from the literal translation by the Fathers of the English Dominican Province (London: Burns Oates & Washbourne, 1923).


———. "Identification and Externality," in The Identity of Persons, ed. Amelie Rorty
University of California Press, 1977); reprinted in The Importance of What We Care About: 58-68.


———. "The Importance of What We Care About," Synthese, 53.2 (1982); reprinted in The Importance of What We Care About: 80-94


———. "Rationality and the Unthinkable," new in The Importance of What We Care About: 177-190.

———. The Importance of What We Care About (Cambridge University Press, 1988).


———. "Fichtes ursprüngliche Einsicht" (Klostermann, 1967); translated as "Fichte’s Original Insight," in *Contemporary German Philosophy*, vol.1 (Penn State University Press, 1982), 15-53.


———. "The Task and Significance of the Logical Investigations*, tr. J.N. Mohanty, reprinted from *Phänomenologische Psychologie (Husserliana* vol. IX).


———. *Critique of Pure Reason*, Unabridged Edition, tr. Norman Kemp Smith (St. Martin’s Press, 1965). All references to this text are given by the standard Prussian Academy pages.


———. _Outside the Subject_, tr. Michael B. Smith (Stanford University Press, 1993).


Lewis, C.S. _The Abolition of Man: How Education Develops Man’s Sense of Morality_


McInerny, Ralph. *St. Thomas Aquinas* (University of Notre Dame Press, 1982).


Meyerson, Denise. "When are my Actions Due to Me?", *Analysis* 54.3 (July, 1994): 171-174.


———. *The View from Nowhere* (Oxford University Press, 1986).


———. *Heidegger and the Philosophy of Mind* (Yale University Press, 1987).


Scheler, Max. *Formalism in Ethics and Non-Formal Ethics of Values: A New Attempt toward the Foundation of an Ethical Personalism*, tr. Manfred S. Frings and Roger L.


———. *Tradition(s): On Community, Rememberance, and Virtue in Classical German Thought* (Indiana University Press, 1997).


Williams, Clifford. Free Will and Determinism: A Dialogue (Hackett, 1980).


