

DEVELOPMENT AND DEFENSE OF A DESIRE-SATISFACTION CONCEPTION
OF WELL-BEING

By

ANTON TUPA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

Copyright 2006

by

Anton Tupa

This document is dedicated to my parents, George and Lonnie Tupa.

ACKNOWLEDGMENTS

I would like to thank Sharon DiFino, David Sobel, Crystal Thorpe, and Jon Tresan for providing helpful comments to this work in its developmental stages. I would also like to thank the audiences of the Ohio Philosophical Association and Florida Philosophical Association for giving helpful comments on parts of two chapters. To Peter Barry, I owe an enormous debt of gratitude for reading over many drafts of my chapters and providing a great deal of helpful advice over the years. Lastly, I would like to thank my advisor, David Copp, for helping me throughout the entire dissertation project and providing guidance at many very difficult times. I am sure that I have left out a number of people who have helped me along the way, and for that, I beg their forgiveness.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	iv
ABSTRACT.....	viii
CHAPTER	
1 THE CONCEPT OF WELL-BEING.....	1
1.1 Introduction.....	1
1.2 Well-Being and Moral Value.....	7
1.3 Well-Being: Narrow and Broad Concepts.....	14
1.4 Well-Being and Self-Regarding Interests.....	17
1.5 Well-Being and Aesthetic Value	20
1.6 Conclusion	34
2 CONCEPTIONS OF WELL-BEING	36
2.1 Four Conceptions.....	37
2.1.1. Mental State Conceptions.....	37
2.1.2 Desire Satisfaction Views.....	41
2.1.3 Explanatory Objective Theories	44
2.1.4 Pluralistic Theories of Well-Being.....	48
2.1.5 Conclusion.....	50
2.2 Critical Discussion of Conceptions	51
2.2.1 The Story	51
2.2.2 The Application of Explanatory Objectivism to the Story	53
2.2.3 Application of Happiness to the Story.....	64
2.2.4 Pluralistic Conceptions of Well-Being.....	70
2.2.5 Conclusion.....	80
3 DEVELOPING A DESIRE-SATISFACTION CONCEPTION OF WELL-BEING	82
3.1 Actual Desire Accounts Critically Discussed.....	82
3.1.1 Some Introductory Remarks.....	82
3.1.2 Intrinsic and Extrinsic Desires.....	83
3.1.3 Problems for Total Desire Satisfaction Accounts	85
3.1.4 Intrinsic Desires.....	87

3.1.5 A Critical Discussion of Actual Intrinsic Desire Views of Well-Being.....	89
3.2 Counterfactual Desire Accounts Developed.....	98
3.2.1 Information and the Relevant Desire Set.....	98
3.2.2 Ideal Advisor Views.....	104
3.2.3 What the Ideal Advisor Advises (or what advice we should listen to)	106
4 IDEAL ADVISOR VIEWS.....	111
4.1 The Internalism Requirement.....	111
4.2 Rosati's Specific Internalism Requirement.....	117
4.3 Full Information and the Conditional Fallacy.....	133
4.4 Knowledge That and Knowledge What Something is Like.....	139
4.5 Propositional and Non-Propositional Knowledge Revisited.....	150
4.6 Railton's Account of Personality.....	154
4.7 Conclusion.....	156
5 TURNING THE DESIRE-SATISFACTION ACCOUNT DEVELOPED ABOVE INTO A THEORY OF PRUDENTIAL WELL-BEING.....	158
5.1 Self-Regarding Interests.....	159
5.1.1 Conceptions of Prudential Well-Being and Self-Regarding Interests.....	159
5.1.2 Self-Regarding Desires: The Problem of Self-Sacrifice.....	160
5.1.3 The Problem Restated.....	162
5.1.4 Overvold's Proposals.....	163
5.1.5 A Problem with the Derived-Desire Interpretation of Overvold's Second Proposal.....	171
5.1.6 Self-Regarding Desires.....	172
5.1.7 Self-Regarding Interests Concluded.....	180
5.2 Non-Moral and Non-Aesthetic Interests.....	182
5.2.1 The Problem.....	182
5.2.2 Moral Desires and the Relevant Desire Set.....	185
5.2.3 Another Approach.....	186
5.2.4 Aesthetic Values and Prudential Well-Being.....	191
5.3 Perverse Advice.....	193
5.3.1 Caring and Well-Being.....	193
5.3.2 Darwall on Caring and Welfare.....	195
5.3.3 Natural Kinds and Definitions.....	198
5.3.4 Sympathetic Concern and Well-Wishing.....	199
5.3.5 The Object of Care and Well-Wishing and "One Thought Too Many"..	200
5.3.5 Perverse Advice and Caring.....	204
5.4 Conclusion.....	206
6 MIXED THEORIES AND DESIRE-SATISFACTION CONCEPTIONS OF WELL-BEING.....	208
6.1 Mixed Theories: What They Are and What Motivates Them.....	209
6.2 Mixed Theories that Involve Desire-Satisfaction Accounts.....	210

6.2.1 Desire/Objective Mixed Theories.....	212
6.2.2 Desire/Mental State Mixed Theories.....	220
6.3 Mixed Theories that Include Mental States	225
6.3.1 Sumner's Mixed Happiness Theory	225
6.3.2 Mental State Accounts Mixed with Objective Elements.....	240
6.4 Conclusion	247
7 DO POSTHUMOUS EVENTS ALTER THE LEVEL OF WELL-BEING ONE HAD WHEN ALIVE?.....	248
7.1 Introduction.....	248
7.2 Overvold's and Hooker's Proposals	249
7.3 A New Beginning	253
7.4 Concluding Remarks	263
LIST OF REFERENCES	271
BIOGRAPHICAL SKETCH	275

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

DEVELOPMENT AND DEFENSE OF A DESIRE-SATISFACTION CONCEPTION
OF WELL-BEING

By

Anton Tupa

May 2006

Chair: David Copp
Major Department: Philosophy

In my dissertation, I develop what I take to be the best version of desire-satisfaction theories of well-being and I defend my favored version against competing theories of well-being.

Desire-satisfaction theories, roughly, are those according to which one's well-being varies with the extent to which some or all of one's desires are satisfied. A desire is satisfied if the desired state of affairs obtains. My favored version of desire-satisfaction theories is quite complex, but the basic idea is that the desires that are true to one's personality are the desires that are relevant to one's well-being.

The first chapter is dedicated to a clarification of the concept of well-being. I settle on the following explication of this concept: a person's well-being is or consists in the non-moral, non-aesthetic, self-regarding interests of the person.

The second chapter of my dissertation is a defense of desire-satisfaction conceptions of well-being against its traditional competitors: mental state conceptions, perfectionist conceptions, and objective list conceptions.

The third and fourth chapters involve developing and defending my favored desire-satisfaction view. The fifth chapter consists of curtailing the desire-satisfaction view, thus far developed, to the conditions of well-being detailed in the first chapter. Thus far, nothing has restricted the desire-satisfaction view to the desires which are not influenced by moral concerns, aesthetic concerns, or desires which are not sufficiently self-regarding.

The sixth chapter is dedicated to defending my preferred desire-satisfaction view against more sophisticated conceptions of well-being.

The final chapter is dedicated to an application of desire-satisfaction accounts of well-being to the issue of whether events that happen after one dies can alter the overall amount of well-being the person had in his or her life as a whole. I argue that events that happen after one dies cannot alter the overall amount of well-being one had in one's life as a whole, not, at least, on desire-satisfaction conceptions of well-being.

CHAPTER 1 THE CONCEPT OF WELL-BEING

1.1 Introduction

My ultimate conclusion is that a desire-satisfaction conception is the best conception of well-being. However, before we get into arguments about whether this is so, first we must have at least a general idea of what the concept of well-being is. I believe that there is a significant amount of misunderstanding about the concept of well-being. This first chapter is an attempt to make clear, at least in a general sense, what this concept is.

Some authors who write on philosophical accounts of well-being give putative synonyms for “well-being” such as “welfare,” “prudence,” “self-interest,” “a good life,” “individual good” and so on. None of these synonyms provide any clearer idea of what the concept of well-being is than any other. Although these synonyms are cue words for people who are already familiar with the topic (and so serve an important purpose in virtue of that), none of them provide a *clarification* of the concept of well-being. What I am looking for in this first chapter is a clarification of the concept of well-being.

Above, I use the terms “concept” and “conception” loosely; I will now give a more precise, semi-technical, explanation of those terms. This will help in two ways: (A) it will help to make clear how fine-grained an account of well-being we should expect at each point in this dissertation, and (B) it will help me in providing a framework for debates over the various accounts of well-being.

Rawls' distinction in *A Theory of Justice* should help here. Early in that book, Rawls distinguishes between a concept of justice and a conception of justice (Rawls, 5-6).¹ Here is what Rawls says about concepts and conceptions of justice: "Thus it seems natural to think of the concept of justice as distinct from the various conceptions of justice and as being specified by the role which these different sets of principles, these different conceptions, have in common" (Ibid, pg. 5). Rawls has an account of the common role of conceptions of justice. Here is one statement of that role: "Those who hold different conceptions of justice can, then, still agree that institutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties and when the rules determine a proper balance between competing claims to the advantages of social life" (Ibid, pg. 5). If one were to generate a conception of how to organize an institution that (1) forbids arbitrary distinctions between people and (2) provides a solution to the claims of different people regarding the social goods, then that conception would be a conception of justice, according to Rawls. More carefully, it would be a conception of the concept of justice that Rawls is trying to explicate. There could be other concepts of justice than justice in the arrangement of social institutions, which is the concept of justice Rawls is after, such as justice in punishment or justice in family life; however, Rawls' project is not to elucidate those concepts of justice.

I think that the distinction between concepts and conceptions can helpfully be applied in my project. Earlier I mentioned that the account of well-being I favor is a kind of desire-satisfaction theory. This theory is, very roughly, that one's life goes well to the extent to which one's desires are satisfied. There are other conceptions. Another is that

¹ Rawls claims to have been influenced by H.L.A. Hart in his *The Concept of Law*.

one's life goes well to the extent to which one is happy and not sad. Yet another conception is that one's life goes well to the extent to which one develops and exercises human nature. These conceptions have in common that they can serve as accounts of a single concept—the concept of prudential well-being. This idea leads naturally to the question, what is the concept of prudential well-being?

On the concept of well-being I am after in this dissertation, *well-being consists in the satisfaction of the non-moral, non-aesthetic, and self-regarding interests of the individual*. As we will see, there are other concepts of well-being—other concepts that could be described as concepts of “well-being.” For this reason, to sharpen my discussion, I will often use the term “prudential well-being” in place of unmodified “well-being.” I shall not be trying to say what “prudential well-being” means in ordinary English. It is rather that I am trying to analyze a concept and have to call it something: I think “prudential well-being” fits the bill at least as well as any other term, but I acknowledge it is not a perfect fit.²

It is important to get clear on the concept we are after because without a clear idea of the concept, we cannot be sure that the various conceptions are competing conceptions of the *same* concept. The goal of the first chapter is to get clear on what the conceptions are *of* before going on to argue for my favored conception. With a clearer understanding of the concept we are after, we can be sure that the debate between the competing

²“Prudential” is used in many different ways, so I should say more about my fairly technical usage. Oftentimes “prudential” is used to describe a course of action or, alternatively, a reason: when it is used in this way, it means a course of action or a reason is suitable given one's goals. With this use of “prudential,” going to the desert would be prudent for someone who wants to live in isolation, all other things being equal. I am using “prudential” in a slightly different sense; I am using “prudential” as a term that captures a kind of value or possible goal—not whether some course of action will obtain the goal or one's reason is a good reason given a goal, but the goal or value itself. With this use of “prudential,” supposing that friendship enhances one's well-being intrinsically, going to the desert is not prudent for the hermit, all other things being equal. It is the second use of “prudential” that I am trying to capture.

conceptions involves genuine disagreements, and not pseudo-disagreements based on confusions about the grounds of the debate.

In an effort to avoid misunderstanding, I would like to say a bit more about prudential well-being in the next few paragraphs. First I will say some more about what it is; then I will say more about why we should study it (what role it plays); lastly I will say more about it to distinguish it from another, somewhat closely related concept. This must all be very brief because I will say a lot more about the first topic throughout the dissertation, and very detailed discussions of the second and third topics are beyond the scope of the dissertation.

In Plato's myth of the Ring of Gyges, we are to imagine that Gyges finds a ring that can make him invisible. Once he puts it on, he finds out very quickly that he can get away with just about anything he wants. His plans are grand: using the ring, he seduces the queen and overthrows the king. These actions are morally wrong, but they seem clearly to serve Gyges' self-interest. Gyges enhances his prudential well-being, I would say. The concept of prudential well-being that concerns me in this dissertation is the concept of the kind of value with respect to which Gyges' life improves in the myth. My proposal is, then, that Gyges' life improves in that he is able, by using the ring, to satisfy more of his non-moral, non-aesthetic and self-regarding interests.

Now I need to say something about why the concept of prudential well-being is worth studying and what role it plays in various contexts. One can agree with me in thinking that there is a concept of prudential well-being and even have an informal understanding in what such well-being consists—and yet wonder why the concept is worth studying.

First, I think that the concept plays a role in our ordinary, pre-theoretic, evaluations of lives. The discussion of the life of Gyges (and the life of the Count of Monte Cristo, which will be my principal example) is meant to illustrate this ordinary, pre-theoretic, notion and how it can be used to evaluate how a life is going. A rigorously philosophical account of prudential well-being could make judgments about people's lives more articulated.

Second, prudential well-being plays a role in many moral and political theories. This may sound paradoxical given that prudential value is restricted to the satisfaction of non-moral interests, but it is not as paradoxical as it may seem. A standard Utilitarian view is that an act is right if and only if it maximizes utility impartially in the long run. Some Utilitarians take "utility" to refer to prudential well-being. So the Utilitarian theory, on this account, is that the right act is that which maximizes prudential well-being impartially in the long run. In political theory, egalitarians sometimes advocate equality of well-being, and I believe at least some egalitarians have in mind prudential well-being. An egalitarian might add that a proper role of the state is to promote the equality of the prudential well-being of its members. There are numerous other moral and political theories that make reference to prudential well-being.

One can agree with me that the concept of prudential well-being plays a role in various ordinary, everyday judgments of lives and also that the concept plays a role in various moral and political theories—and yet think "so much the worse for those judgments and theories." However, such a view may be based in a confusion between prudential well-being and selfishness.

The term “selfishness” does not appropriately capture the concept I am after. Someone behaves selfishly, roughly, when she disregards the welfare of others when concern for the welfare of others is appropriate.³ The basic idea is that selfish behavior involves neglect of the welfare of others. Self-interested behavior, in contrast, or behavior motivated by concern for the agent’s well-being does not necessarily include such neglect. Self-interested behavior can include regard for the well-being of others.⁴ If one conflates the concepts of self-interest and selfishness, then any moral or political theory that is based on the concept of self-interest may look extremely implausible. But such conflation is a mistake—the two concepts are different. Thus, it would be a mistake to dismiss our concern with exploring the concept of prudential well-being on the basis of a conflation of actions done out of a concern for prudential well-being with selfishness.

In this brief interlude on prudential well-being, I have tried to give some motivation for thinking that there is such a concept and that the concept plays a role in our ordinary value judgments. Additionally, I have tried to give a few examples of moral and political theories in which the concept of prudential well-being plays a role. I also stressed the importance of distinguishing between selfish behavior and behavior motivated by prudential well-being.

Might there not be some concept of well-being, besides prudential well-being, worthy of study? I think that there is. The situation I face with the concept of well-being is similar to the situation Rawls faces with concepts of justice. Rawls chooses to study

³ Strictly speaking, this is only a necessary condition of selfish behavior. That condition should be enough to make the difference between the concept of prudential well-being and selfishness clear.

⁴ Suppose that a person is faced with only two possible courses of action; both would enhance her well-being to the same degree. Suppose furthermore that the first possible course of action increases the well-being of other people while the second does not. The courses of action are equally self-interested. Yet the second is selfish and the first is not.

the concept of justice in the arrangement of social institutions and I choose to study prudential well-being: these are valid choices, despite the fact that there are other, closely related, concepts worthy of study. The common role of the various conceptions of prudential well-being in a diverse array of contexts is what makes me think the concept ripe for critical examination.

In the remainder of Chapter 1, I will explain my proposed account of the *concept* of prudential well-being. Later, in Chapters 3-5, I will go on to explicate and develop a desire-satisfaction *conception* of well-being.

Recall that I claim that the concept of prudential well-being is that well-being consists in the satisfaction of the non-moral, non-aesthetic, and self-regarding interests of the individual. That makes for three key conditions on what constitutes an interest of the proper sort: It must be (1) non-moral, (2) non-aesthetic, and (3) self-regarding.

Heretofore, I have not presented any reason for believing those three conditions are correct; indeed, I have not explained what those conditions are in detail. In the remainder of this chapter, I take up each of these three conditions in turn, moving from the non-moral condition to the requirement that the interests be self-regarding and then to the requirement that the interests be non-aesthetic. I will introduce each of these conditions using well-known fictional stories, and then I will go on to discuss the conditions in a more rigorous philosophical manner.

1.2 Well-Being and Moral Value

It is helpful to begin with a fairly in-depth case study that will illustrate the concept of well-being. My case study will be the life of the protagonist in Alexandre Dumas' novel *The Count of Monte Cristo*. Let me give those readers unfamiliar with the story a very brief outline of the plot. As the novel begins, the protagonist, Edmond Dantes, is

living quite well. Unbeknownst to him, there is a plot against him and he is thrown in jail without any semblance of a fair trial and without chance of seeing or speaking to anyone. Eventually he escapes and discovers a great treasure. At this point he becomes known as The Count of Monte Cristo. He uses this treasure for retribution against those who plotted against him and to benefit those who truly loved him. In this phase, Monte Cristo sees himself as a “divine angel of retribution.” In the end, but after he has accomplished his goals, he realizes the emptiness of a life of vengeance and moves on to a different path in life.

The plot is fantastically complex and serpentine. I do not have the space to give more detail about its general structure.⁵ Nevertheless, the changing fortunes of Dantes’ life should provide stories that are helpful.

As we begin the novel, Edmond enjoys a high degree of well-being. He is illiterate and does not have a formal education. But he is a good son, is about to be promoted to the rank of captain on a ship, and is engaged to be married. His life, intuitively, has a high level of well-being. The plot develops and several people, each with motives of his own, conspire against Edmond. They succeed and Edmond is thrown in an island prison called the Chateau d’If. Edmond never receives much of an explanation as to why he is in prison. The worst kinds of conditions imaginable are standard fare in this prison. He is isolated in a small dungeon with very little light, he cannot communicate with anybody and he has nothing to do. He is more than just bored; his poor conditions and inexplicable imprisonment almost drive him insane. Edmond’s change of fate is an easy case for the

⁵ Just to give the reader a hint, the planning and bringing about (in detail) of Napoleon’s return to The Continent from exile on the Isle of Elba is but a mere sub-plot in this novel.

concept of well-being: his life is clearly going poorly. His most basic needs (where these are understood as food, clothing and shelter) are met, but surely, this is not enough.

A while after Edmond enters the Chateau d'If, another prisoner named Abbe Faria tunnels into Edmond's prison cell. The Abbe is trying to tunnel his way out of the prison but ends up in Edmond's cell—now the two cells are connected by the tunnel, but the tunnel does not go outside the walls of the prison. The Abbe is a learned person who has several books (perhaps they are better called “manuscripts”) in his cell and quite a few contraptions that help make life go better. He begins to teach Edmond to read and then goes on to teach him various languages and several subjects. Lastly, the Abbe, on his deathbed, tells Edmond of a great treasure on the small uninhabited island of Monte Cristo. The Abbe dies and Edmond escapes. Edmond's life has been improving; first he has someone to talk to; then he develops his talents and he starts to live the kind of life he wants to live again. Eventually, he gets his freedom.

But now let us look at some more difficult cases. As the novel goes along, Edmond's conditions improve in various ways. Let us look to see if these improvements are improvements in well-being.

Edmond goes to Monte Cristo and finds the great, almost unfathomably large, treasure. Because he does not want people to know that he has escaped from prison, he changes his name to The Count of Monte Cristo—hereafter referred to as “The Count” or “Monte Cristo.” Has the mere fact that he has found this enormous treasure increased his

well-being? No, the wealth will often provide *means* to well-being; it does not provide well-being itself.⁶ So the concept of well-being cannot be reduced to wealth.

Economists will often look at wealth when trying to measure well-being, and if not exactly well-being, perhaps utility, which is a concept that is a nearly-related cousin.⁷ Although wealth may be thought of as the possession of money and goods, perhaps our idea of wealth should also include the ability to spend money and to use and have access to possessions. So, wealth (as we should think of it) may be a slightly richer notion than just having money and possessions. Still, this richer notion of wealth is not the same as well-being. It would be a mistake to *identify* wealth with well-being, for one can be wealthy and have a low level of well-being, and one can be poor and yet have a high level of well-being. Perhaps, however, economists intend merely to treat wealth as a sign of well-being. If so, I do not have a *philosophical* disagreement with them, but I have two brief comments. The first is that I am not concerned with finding the signs of well-being; I am concerned with finding in what well-being consists. Second, many economists are moving away from wealth as a sign of well-being.⁸ To bring this discussion back to *The Count of Monte Cristo*, there is a correlation between the Count finding the great treasure and his life going better, but the treasure is only a means to his

⁶ I am in concert with Aristotle on this subject. He is looking for an analysis of a similar concept when he says: “wealth is evidently not the good we are seeking; for it is merely useful and for the sake of something else” (*EN*, 1096a6-8).

⁷ “Utility” as economists use the term, may be after something slightly different than “well-being” as I understand it.

⁸ Both the World Bank and the UN publish, at regular intervals, world-wide data that are far better signs of well-being (such as literacy rates and broader education data, employment data such as the kinds of jobs people can have, and so on).

increasing his well-being. The important point is that opulence cannot be identified with well-being.

As I mentioned above, after securing the treasure, Monte Cristo has two goals: to get revenge on those who plotted against him, and to benefit those who truly cared for him. I will discuss each of these projects in turn, for they raise interesting issues for our discussion of well-being. First I will discuss the Count's revenge and the possibility of placing a moral condition on well-being. The Count is willing to go to great extremes to get his revenge. He does not seem willing to "do whatever it takes," but some of his actions are, at best, morally inappropriate. At one point, when The Count is trying to bankrupt one of his wealthy adversaries, he manipulates a kind of commodities exchange (much like the modern day stock market). Many people lose fortunes due to the Count's actions (the Count had to have foreseen this), but the Count achieves his goal of ruining the one person he was after.⁹

In another instance, the Count gives the wife of one of his adversaries information about poisons and ways of poisoning people without being detected. The Count then goes on to give her the poison itself. Monte Cristo does not give her the end or goal (as if by manipulation or command), she provides that herself. She then proceeds to poison no fewer than six people—killing five of them, including herself and her son. Monte Cristo knew of her poisonings, which took place over a number of weeks, and did not stop

⁹ As is always the case in Dumas' novel, the story is not quite so simple. The man Monte Cristo is trying to ruin, Danglar, loses great deal of his fortune in the stock market. Danglar then steals a fortune from a hospital and runs away. Danglar is then captured and imprisoned. When in his cell, he must pay exorbitant sums of money in order to get any food, thereby bankrupting him, even of his stolen riches. The lesson that Danglar is to learn is that life is more precious than money (Danglar originally conspired against Monte Cristo for financial gain).

them.¹⁰ Perhaps, in some extremely technical sense, the Count's hands were clean. But, we must keep in mind that he knowingly contributed to the deaths of many innocent people. Additionally, the Count conspired to have Villefort's daughter marry her half-brother (the blood relation of these two individuals unbeknownst to everyone save Monte Cristo, and perhaps his servants).¹¹ There is surely something wrong in that.

There are other instances of like actions that I will not go into. Monte Cristo does not seem to be willing *merely* to ruin the lives of those who conspired to ruin his. For each of the four who conspire against him, Monte Cristo has a particular lesson to teach them before they are murdered, commit suicide, or are ruined. Furthermore, Monte Cristo reveals his true identity to each of them at the time of ruining their lives, perhaps so that they will learn their lesson better. Despite this moral education that Monte Cristo is doling out, and despite the fact that he does not seem to be willing to go quite so far as to do "whatever it takes," some of his actions are wrong, to say the least.

The question we must deal with now is whether the achievement of his plans and projects, expending and developing of his natural talents towards these immoral ends, and his pleasure in these projects and results could count as contributing to Monte Cristo's well-being. The Count is aimed at, or at least involved in, dubious moral enterprises.

Should the above-mentioned actions be *excluded* from increasing the Count's well-being

¹⁰ Villefort's wife (his second), Madame de Villefort, tries to poison all of those people who stand in the way of making her son inherit several fortunes. Her son with Villefort does not stand to inherit any money because Villefort has a daughter by his first marriage. The people Madame de Villefort poisons are: (1) a husband and (2) wife who have a fortune who are relatives on her husband's side, (3) she tries to poison her father in law, who also has a fortune, but instead kills his servant, (4) she poisons, but does not kill, her daughter in law. At this point she is found out and she poisons and kills (5) herself and (6) her son. Villefort originally conspired against Monte Cristo to save his own reputation. This string of poisonings, in addition to several other events, ruins Villefort's reputation.

¹¹ The details of this plot are too complex to go into.

in virtue of their immorality? My answer is “no.” Now, I do not wish to deny that there is a moral dimension to a life’s going well (without restriction), but I am interested in the specific notion of prudential well-being.

As I have said, I am willing to take it for granted that a life can go better or worse as evaluated by moral criteria. However, evaluating lives in light of moral criteria is not the topic of this dissertation. As I will show later in this chapter, some authors who use terms such as “well-being,” “welfare,” “individual good,” etc. are writing about a concept that is essentially moral. It is important to be clear, at this early stage, that I am after a different concept.

Here is, I think, a helpful way of putting the issue: we can interpret “well-being” in a narrow and a broad sense. In the broad sense, the “well” of “well-being” is a general and overarching term of evaluation that applies to one’s being or life. If a life goes well in this broad sense, then the life is more choiceworthy without restriction. In the narrower sense, “well-being” refers to just a single dimension of choiceworthiness—the prudential dimension, as I call it.

Returning to the Count’s life, my contention is that he can have a higher level of prudential well-being when he succeeds in his projects, develops and uses his talents, and is pleased in his pursuits and successes, even when his projects are immoral. So, when the Count manipulates the commodities exchange and thereby bankrupts his adversary (also bankrupting people completely uninvolved in the conspiracy against the Count), or when he gives poison to the wife of another adversary knowing full well that she will poison innocent people, the Count’s life goes prudentially better. He enhances his prudential well-being.

It is also possible for a project to be morally right and increase the agent's prudential well-being. It is a confusion to think that morally wrong actions always increase well-being. It is equally a confusion to think that morally right actions always decrease well-being (or do not contribute to well-being). It is just that to enhance a life morally is not necessarily to enhance it with respect to prudential well-being, nor vice versa. So although I illustrate the distinction between moral and prudential well-being with a discussion of the Count's immoral actions, please do not read too much into this.

1.3 Well-Being: Narrow and Broad Concepts

I would like to examine two fairly recent works where I think philosophers might seem to be after the same concept as I am (the narrow concept of prudential well-being), but are in reality, after the broader concept of unrestricted goodness of a life. David Brink, in the final chapter of his book *Moral Realism and the Foundations of Ethics* examines various proposals for theories of value. He uses terms such as “well-being” and “welfare” in addition to “value.” Brink mentions a case of a Nazi who has plans to persecute Jews and claims that the Nazi's accomplishment of his goals in this example could not count to increase the value of that life (Brink, pg. 227). The particular conception of well-being Brink settles on leaves room for satisfaction of “reasonable” and “admissible” projects as enhancing one's well-being (Ibid, pg. 233).¹² Given what Brink says, and the example of the Nazi that he uses, it seems that Brink places a moral requirement on what can count as one's good.

¹² Brink's final and full-bodied version of well-being is decidedly Aristotelian. But I am not at this moment concerned with which conception of well-being is best. I am here trying to get a fix on the concept of well-being and be clear about which concept I am after. It is how Brink argues against certain conceptions of well-being that tells us what he takes the concept to be.

We have had to get to Brink's concept of well-being *indirectly*: that is, we have had to figure out what concept he was after by looking at how he argues for his favored conception. The next author I will discuss speaks more *directly* about well-being at the level of concepts—at the level of generality of concepts as opposed to conceptions.

Stephen Darwall, in his recent book *Welfare and Rational Care* offers an analysis of welfare. He says: "What is for someone's good or welfare is what one ought to desire and promote insofar as one cares for him" (Darwall, pg. 7, italics omitted). Darwall uses all of the regular synonyms for welfare such as "a person's good, interest, well-being, or welfare" (Ibid, pg. 1) and even "prudential value" (Ibid, pg. 12). Given this, one might think that he is after the same concept I am, however, I intend to show that he is after a different concept—which turns out to be much like the concept that Brink is after.

The "ought" that appears in Darwall's analysis of welfare is not obviously moral. If it were, I would well be on my way to showing that Darwall has in mind a concept just like Brink's. Unfortunately, things are not that easy. Elsewhere, Darwall replaces "ought" with "should" (Ibid, pg. 8) and "would rationally" (Ibid, pg. 9). I do not think that Darwall is using "rational" in a moral sense. Darwall gives a general idea of what he means by "rational" by allowing "makes sense, is warranted or justified" to replace it (Ibid, pg. 9).

Another important issue is that Darwall's analysis is counterfactual in nature. The "insofar as one cares for him" phrase is important here. Darwall means that a person's welfare consists in what one *would* rationally desire for the person *were one to care for him*. We could even go so far as to give a "possible worlds" analysis of this: What is in

someone's welfare is what one rationally wants for him in the nearest possible world in which one were to care for him.¹³

Now that I have explained Darwall's analysis, I can examine whether or not he is after the same concept I am. Darwall's claim is that something, x , contributes to someone's, A 's, welfare if and only if x is what someone, B , would rationally want for A , were B to care for A , other things being equal. Nothing excludes A and B from being the same person (Ibid, pg. 20). Now, as I understand this, Darwall is searching for an account of something akin to, or even identical to, the good life. I do not see any elements of his formulation that limit the range of choices to prudential value in the narrower sense described above. The morally good life, prudentially good life, and perhaps even the aesthetically good life, all could be elements of the kind of life one would rationally want for someone, were one to care for that someone.

Darwall, like Brink, ultimately chooses an Aristotelian conception of well-being. Brink, Darwall, and myself use many of the same terms for what we are after: "welfare," "well-being," "self-interest," "prudential value," and so on. However, if I am right, we are after very different concepts. There are broad and narrow concepts of well-being: the broad concept is similar to the concept of the good life, the narrower concept excludes at least moral value and focuses only on prudential value. Further explanation of what this narrower value consists in must wait. They are after a broader concept and I am after the narrower concept.

I do *not* claim that Brink and Darwall are wrong in any serious way. I do take myself to have made the difference between us clear: we are after different concepts

¹³ I am uncertain on whether "insofar as" in Darwall's usage has any implications about the *degree* of care. The degree of caring could be relevant to Darwall's account, but the text is not clear on that issue.

despite the fact that we use the same terms. I take “prudential well-being” to denote the concept of well-being I am after in this dissertation—this is contrary to the usage of Darwall and Brink, but I think that “prudential” is slightly more apt to the narrower concept that I am trying to explicate than it is to the wider concept that is of interest to Darwall and Brink.

Let us return briefly to the story of the Count. The Count’s revenge then, even though extreme and immoral, would still count as increasing his well-being. At least, it would not be excluded in virtue of its being immoral. Something similar should be said about projects which are morally good. Something should not be included in the list of what makes someone’s life go well (in the sense I am after), merely in virtue of its moral goodness. There is a great deal more that must be explained about the concept I am after, and for this, we must continue with *The Count of Monte Cristo*.

1.4 Well-Being and Self-Regarding Interests

As I mentioned above, the Count directs his immense wealth and talents towards a pair of purposes—heaping benefits on his true friends and seeking revenge on those who plotted against him. I have already discussed the relevance of the motive of revenge, but what of the motive of heaping benefits on those who really did care for him? Although the bulk of the novel is dedicated to the Count’s revenge, the first thing the Count does with his newly found wealth and secret identity is to benefit those who truly loved him. His mother died before he went to prison, his father died while he was in prison, his fiancée married a man who (unbeknownst to her) was one of those who originally plotted against him, but his former employer, Monsieur Morrel, the owner of the ship Monte Cristo was to captain, is still alive and much in need of help. Morrel is in debt and cannot afford to pay the debt back. Furthermore, Morrel’s reputation is at stake. The Count

risks the loss of his secret identity to help Morrel out. Monte Cristo then gives Morrel money to pay off his debts.¹⁴ While we may not view the Count's gift of the money as a sacrifice (given that he has so much of it), the danger of people finding out who he is and where his fortune came from are real and significant threats to his well-being.

Is the success of Monte Cristo's plan to save Morrel an increase in his (Monte Cristo's) well-being? In some broad sense, clearly the Count gets what is in his interests, but this may be a case where the interests are not sufficiently self-regarding to count as prudential. As I interpret the story, the Count's actions are self-sacrificial. They do not enhance the Count's well-being, even though they serve the count's goals. To understand self-sacrificial acts, we need to distinguish between self-regarding and other-regarding interests. I will discuss this distinction in detail in a later chapter. For now, however, notice that, intuitively, self-sacrificial behavior may lead to the satisfaction of other-regarding interests but normally not the satisfaction of self-regarding interests. Indeed, with acts of self-sacrifice, there is setback (at least as foreseen by the self-sacrificing agent) to self-regarding interests.

There are those who might deny that there are genuine acts of self-sacrifice. Let us call the theory that everyone always acts in her self-interest "Psychological Egoism." Psychological Egoists must deny that there are any real acts of self-sacrifice. As I stated above, the interests, the satisfaction of which enhance prudential well-being, must be self-regarding. If Psychological Egoism were true, then the requirement that the interests

¹⁴ In what should now be a familiar dramatic fashion, the details of the story are more complex. Morrel, who takes his honor very seriously, determines that he will kill himself at the moment just before the debt is due—thereby saving the honor of his family. The Count, in a secret identity (I suppose this would be a "second-level" secret identity), as a representative of a banking house, purchases all of Morrel's outstanding debts to other banking houses and then forgives the entire sum. Of course, all of this happens just a moment before Morrel attempts to commits suicide.

be self-regarding would be vacuous or empty in a strong sense. This is so because, according to Psychological Egoism, all actions are motivated by self-regard, and thus no interests are excluded from being relevant to one's prudential well-being.¹⁵ Now, I think that Psychological Egoism is *false* and I will argue against it in a later chapter of this dissertation.¹⁶ But for now, let us *assume* Psychological Egoism is false.

Many philosophers have recently dealt with the distinction between “self-regarding” and “other-regarding” interests, though they all use different terms in drawing the distinction.¹⁷ The basic idea is that there must be some restriction placed on which interests are such that their satisfaction counts as enhancing the agent's prudential well-being. As I said I will discuss the distinction between self- and other-regarding interests in more detail in a later chapter of this dissertation. At this moment, the important point is that a person's prudential well-being is not necessarily enhanced by the satisfaction of other-regarding interests.

Returning to *The Count of Monte Cristo*, when the Count saves Morrel from his debts, the Count's well-being is not enhanced, so I say. The Count's act is self-sacrificial and results in the satisfaction of his interests, but the interests which are satisfied are other regarding—they concern the welfare of Morrel and not the Count. A detailed

¹⁵ The Psychological Egoist, for all I have said, *could* allow that there are other-regarding interests but nevertheless claim that they never motivate. This is a conceptual possibility, but it does not seem to have much initial plausibility. I will discuss it in a later chapter.

¹⁶ In a way, my project would be simpler if Psychological Egoism were true because the requirement that interests be self-regarding would not require any explanation. However, because I think Psychological Egoism false, I take on the burden of explaining the distinction between self- and other-regarding interests.

¹⁷ Two notable examples are Ronald Dworkin and Thomas Scanlon. Dworkin, in his *Sovereign Virtue*, distinguishes between personal and impersonal preferences for a theory of welfare (pg. 25-28). Scanlon discusses a similar distinction, though using different language, in his *What We Owe to Each Other* (pg. 114-115, 120-124). Both authors claim to be influenced by Parfit's discussion of this topic in his *Reasons and Persons* (pg. 493-502).

explanation of how this is possible, on my preferred conception of well-being, is in a later chapter.

1.5 Well-Being and Aesthetic Value

So far, I have explained why we need to understand prudential well-being as involving the satisfaction of non-moral, and self-regarding interests of the individual. Might there be more restrictions on which interests, when satisfied, increase well-being? I would like to look at another proposal.

I have briefly mentioned above the possibility of regarding certain aesthetic characteristics of a life as enhancing the overall prudential well-being of the person. This idea has not been much discussed. I think it is interesting, however, so I will look at a few accounts of aesthetic value and then address some significant possibilities.

G.E. Moore famously claims that the *appreciation* of beautiful objects is intrinsically valuable. He also makes the stronger and more controversial claim that beautiful objects are intrinsically valuable apart from any appreciation of them.¹⁸ Now, beauty as such, cannot enhance anyone's prudential well-being. That is because the beauty of beautiful objects might not be in a person's life or appropriately related to a person's life. Well-being is necessarily something of value in a person's life. It is not merely an impersonal value

¹⁸ For Moore's arguments by the method of absolute isolation, see §50 and §119-§121 of *Principia Ethica*. It should also be noted that Moore, in a later work *Ethics*, does not mention beauty as intrinsically valuable and even is inclined to adopt a view that rejects the possibility of beauty being intrinsically valuable. In *Ethics*, Moore claims: "it does seem as if nothing can be an intrinsic good unless it contains both some feeling and also some other form of consciousness; and, as we have said before, it seems possible than amongst the feelings contained must always be some amount of pleasure" (Moore, *Ethics*, pg. 107). This requirement of pleasure, of course, eliminates the possibility of objects being intrinsically valuable in virtue of their beauty alone.

What follows from this discussion is that if aesthetic value enhances someone's well-being, it must somehow involve the agent whose well-being is at stake. There seem to me to be three obvious ways to go here: (1) we can look at the beauty of a person, or (2) we can look at the subjective or first-person aesthetic appreciation of beautiful things by an agent and claim that that has something to do with his well-being, or (3) we can look at the aesthetic beauty of the life story of an agent. One thing to keep in mind as I proceed through these three options is that I am not attempting to determine whether or not these ways of involving beauty in one's life are really valuable. I am concerned with whether they could enhance one's well-being in the narrow sense, solely in virtue of their aesthetic character.

I begin with the first idea, the idea that a person's well-being can be enhanced by her beauty. This idea does not have much plausibility as serving as a component of prudential well-being, but an examination should prove helpful. A discussion of Oscar Wilde's work, *A Picture of Dorian Gray*, should illustrate the proposal that is on the table. I will briefly summarize the story. As the novel begins, Dorian Gray is young and extremely handsome. He is, at least to a large extent, innocent and ignorant of his good looks (or at least the extent of them). Basil Halliward, a developing painter, is completing a portrait of Dorian Gray. Halliward is so influenced and inspired by the beauty of Gray that he is convinced that he can develop a new style of painting for the modern age with his portrait. Upon completion, Dorian looks at it and realizes the power of his own striking beauty. Struck by his own beauty, Dorian exclaims:

"How sad it is!" murmured Dorian Gray, with his eyes still fixed upon his own portrait. "How sad it is! I shall grow old, and horrible, and dreadful. But this picture will remain always young. It will never be older than this particular day of June . . . If it were only the other way! If it were I who was to be always young,

and the picture that was to grow old! For that—for that—I would give everything! Yes, there is nothing in the whole world I would not give! I would give my soul for that!” (Wilde, pg. 168)

Something quite magical happens at this exclamation. As Gray eventually finds out, he is to remain forever young while the image of him in the portrait will age over time. Gray keeps his remarkable good looks while the portrait shows the signs of age.

This first proposal does not even seem to be a plausible account of how aesthetic value could enhance a person’s well-being. I have never heard of anyone endorsing it in philosophical writing. The aesthetic value that makes a life go better must go deeper than the mere physical beauty of a person. The story of Dorian Gray’s life is in fact supposed to illustrate that something starts to go very bad for him at the time of the magical transformation; that his life goes worse, not better, as he continues to live as a beautiful person physically. Could prudential well-being be associated with this kind of beauty? On this account, Gray’s well-being would forever be enhanced by his beauty. This is surely implausible. I am certainly willing to grant that, throughout history, beauty has been instrumental to prudential well-being, but it simply cannot be an element in a plausible account of the of prudential well-being.¹⁹ As we will see below with the third proposal, there is another way of thinking about the beauty of a life that serves as a far more plausible candidate for being a component of prudential value.

What of the second proposal, the idea that a person’s prudential well-being can be enhanced by aesthetic experience and appreciation? The third major character of *The Picture of Dorian Gray* is Lord Henry Wotton. He is a friend of Basil’s and turns out to

¹⁹ Merely being beautiful is not even plausibly thought of as constituting some broader concept of well-being, if there is such a value. The case of Dorian Gray illustrates this in that Dorian is a morally wicked person and yet remains beautiful.

have a profound influence upon Dorian Gray. Harry, as he is called, expounds his theory of how to live in several speeches throughout the novel. Dorian comes to accept and live by Harry's theories; which promote a kind of aestheticism.

Perhaps the most basic and fundamental element of Lord Henry's theory is captured in the following quotation: "The aim of life is self-development. To realize one's nature perfectly—that is what each of us is here for" (Ibid, pg. 158). Lord Henry believes that there is an end of life and that is to develop one's nature. The theory is described twice as a "new Hedonism" (Ibid, pg. 164, 286), though I think Wilde uses the term very loosely. On this note, Lord Henry advises: "Be always searching for new sensations. Be afraid of nothing . . . a new Hedonism—that is what our century wants" (Ibid, pg. 164). We need to be careful in our understanding of what kinds of sensations Lord Henry is talking about. I do not believe that he means just any sensation, for he already limits it to pleasures (understood broadly). But if we look at his disciple, Dorian and his actions, Dorian clearly seeks artistic sensations, specifically, the sensations of beautiful things.²⁰ So although the theory of Lord Henry begins in abstract terms such as "developing one's self," something more like a kind of aestheticism or pursuit of aesthetic experience is what he has in mind.²¹ An additional idea that seems implied by what Lord Henry says is that it is not just aesthetic experiences that are to be pursued, but aesthetic appreciation. There seems to be an "appreciative" element in Lord Henry's

²⁰ Indeed, art is described as "simply a method of procuring extraordinary sensations" (Ibid, pg. 379).

²¹ There was a movement that began in 1880's England and ran through the turn of the century sometimes called 'Aestheticism.' For an excellent, but brief, discussion of this movement and Wilde's role in it, see Stephen Calloway's "Wilde and the Dandyism of the Senses." As Calloway describes the movement, it is based in developing one's aesthetic response to beauty in the world. This, of course, is slightly different from merely having a lot of aesthetic experiences. One explanation for tying together Lord Henry's claims about (1) developing one's self and (2) having a myriad of aesthetic experience, primarily of beauty and involving pleasure, is that in these experiences, Dorian can refine his aesthetic sensibilities.

advice in that he seems to suggest a more cognitive, active, and critical element is necessary for the right kind of aesthetic value, rather than just passively “experiencing” something. Aesthetic appreciation makes for a far more plausible account of something that enhances well-being.

The entirety of chapter XI of the novel is dedicated to the pursuits and experiences of Dorian Gray. Wonderful music and jewelry are at the forefront of what he seeks to experience. However, he also fancies converting to the Catholic Church for the experiences that accompany Catholic worship. Gray also resorts to opium use, though that could be interpreted as consistent with his pursuit of fantastic aesthetic experience. This aestheticism so dominates Dorian Gray that the suicide of someone he once loved doesn’t make him feel sad, but instead, he appreciates the beauty of the tragedy of her life. It would be a grave misunderstanding to think that tragedy is ugly; tragedy is not ugly. Tragedy is among the most beautiful of art forms. The way the story is set up, Dorian, in his pursuit of aesthetic experience, becomes full of vice and lacks all moral virtue.²² This change in moral character is reflected (paradoxically), as increased ugliness and decay in the painting.

This second proposal for a way in which something of aesthetic value could contribute to well-being seems in danger of confusing aesthetic *value* in a life with aesthetic *valuing* in a life. Even if a person had a great deal of aesthetic appreciation, she still could live an ugly life. Indeed, Dorian Gray seems to be such an example of this.

²² The aesthetic movement often eschewed morality. The term “decadent” is often attached to “aesthetic” in describing the movement. There are numerous instances in *The Picture of Dorian Gray* where Lord Henry claims that ethics has no role in aesthetics. On my schema, I would categorize the aesthetic movement as the pursuit of aesthetic value to the deliberate exclusion of moral value—with perhaps a somewhat surprising silence on prudential value (though I am but an amateur historian).

One of the things that makes *The Picture of Dorian Gray* such a thought provoking novel is that Dorian seems to have beauty in his person (the first proposal) and in his experiences (the second proposal) and yet still he fails to be a beautiful person in a clear and strong sense.²³ The second proposal seems either wrong or incomplete as an account of how aesthetic value can enhance well-being, even when “well-being” is understood in the broad sense.

I am willing to grant for the sake of argument that the aesthetic value of a life has something to do with the aesthetic experiences and appreciation of the person who is living the life, though I suspect that many will disagree with that. Still, I do not think that the aesthetic value of a life, so conceived, can plausibly be thought to enhance prudential well-being, in virtue of its aesthetic aspect. The initial plausibility of this proposal has to do with the fact that aesthetic appreciation seems like a specific kind of pleasure. Lord Henry suggests aesthetic appreciation is a kind of pleasure. If indeed he is right, then this second proposal could be accommodated by certain conceptions of prudential well-being, though only very loosely, as I will explain below. It is important to note that, on this reading of aesthetic appreciation, I am *not* saying that aesthetic appreciation does contribute to prudential well-being, but merely that the idea that it does it is not ruled out conceptually.²⁴

Is aesthetic appreciation a kind of pleasure? It may be. J.S. Mill, in the second chapter of his *Utilitarianism*, is famous for distinguishing between higher and lower

²³ Interestingly, if I am right, one of the lessons of the novel is a rejection of the very aestheticism that Oscar Wilde was supposed to endorse; or at least a much more sophisticated account of what is to be pursued according to aestheticism must be provided.

²⁴ Here, I am using “concept” and “conception” in their technical usages from earlier in the chapter.

pleasures. The higher pleasures are supposed to be the intellectual ones while the lower pleasures are supposed to be the bodily ones. If intellectual pleasures such as the pleasures of philosophical contemplation or friendship could count as “higher quality” pleasures, it is not clear why aesthetic appreciation should not also count as intellectual pleasures. Mill thinks that a cultivated mind allows for the happiest sort of person because of the increased potential for higher pleasures. Mill says:

A cultivated mind—I do not mean that of philosopher, but any mind to which the fountains of knowledge have been opened, and which has been taught, in any tolerable degree, to exercise its faculties—finds sources of inexhaustible interest in all that surrounds it; in the objects of nature, the achievements of art, the imaginations of poetry, the incidents of history, the ways of mankind past and present, and their prospects in the future. (Mill, Ch. 2, paragraph 13)

In this quotation, Mill lists several things that increase well-being, including taking an interest in the achievements of art. However, it is important to understand that for Mill, all of the items on his list increase well-being in virtue of their being sorts of pleasure. It is not in virtue of the aesthetic aspect that a mental state, such as aesthetic appreciation, could enhance prudential well-being.

As we have seen above, aesthetic appreciation makes for a better account of something of value within a life that could enhance well-being. But it appears that aesthetic appreciation is a kind of pleasure. After all, one can have an aesthetic *experience* of something disgusting or ugly; there is no requirement that the thing experienced is of positive aesthetic value. Aesthetic appreciation, on the other hand, seems to require that the thing appreciated, at least by the lights of the “appreciator,” be beautiful in some way. We could even set up a new category; call it “aesthetic disapproval.” Aesthetic disapproval, let us say, involves aesthetic experience and

employs the intellectual faculties but its object is taken by the “appreciator” to have negative aesthetic value.

What makes one attitude aesthetic appreciation and another aesthetic disapproval? One very intuitive answer is that aesthetic appreciation is a preferred state of mind while aesthetic disapproval is not preferred. This fits neatly with Mill’s account of what makes a pleasure higher and lower—choice by competent and informed judges (Mill, Ch. 2). Might there be other answers to the question about what makes a state of mind that of aesthetic appreciation as opposed to aesthetic disapproval? There could be, but it is hard to see what they would look like. Pursuing other lines of thought would be highly speculative. I think that the important point is that the more we look at what aesthetic appreciation could consist in, the more it fits with a very popular and influential account of pleasure. Pleasures are not ruled out conceptually as candidates for constituents of well-being: the arguments and debates over the various conceptions of prudential well-being is something I will get into later in this dissertation

I do not think that the second proposal is a plausible account of how well-being of a life can be enhanced (as I think the case of Dorian Gray illustrates). Moreover, it has trouble distinguishing itself from a kind of sophisticated pleasure or desire theory. If aesthetic appreciation is a kind of pleasure then since one possible view of prudential well-being is that it consists in a pleasurable life, we certainly cannot rule out the second proposal conceptually. However, this is not a problem for my proposal about prudential value. I say that only non-aesthetic interests are such that their satisfaction enhances well-being. But this does not exclude pleasure taken in aesthetic objects, so it does not

rule out the second proposal. However, I think that there is a far better proposal for a way in which aesthetic value in a life can enhance well-being, and I turn to it now.

The third, and in my mind, most plausible account as to a way in which something of aesthetic value might enhance well-being is that a kind of narrative structure of a life might do it. Lives have narrative structures. More properly speaking, *life stories* have narrative structures. Let us set aside the difference for now. Some lives have more beautiful narrative structures than others. The proposal to be examined is that the narrative structure of a life enhances or detracts from the prudential well-being of the person living that life. Let me return to the Count's life because, although it is far-fetched and fantastic, it is not mystical and supernatural as Dorian Gray's life is. Even though the Count's life is fiction, it is more naturalistic than is Dorian Gray's.

The Count's life has a certain literary quality to it. Let us say that a life has a "literary quality" if it has a pleasing narrative structure. Is the prudential value of a life always enhanced by events that give it a more pleasing narrative structure? Some philosophers have thought so; possibly Michael Slote and more likely David Velleman.

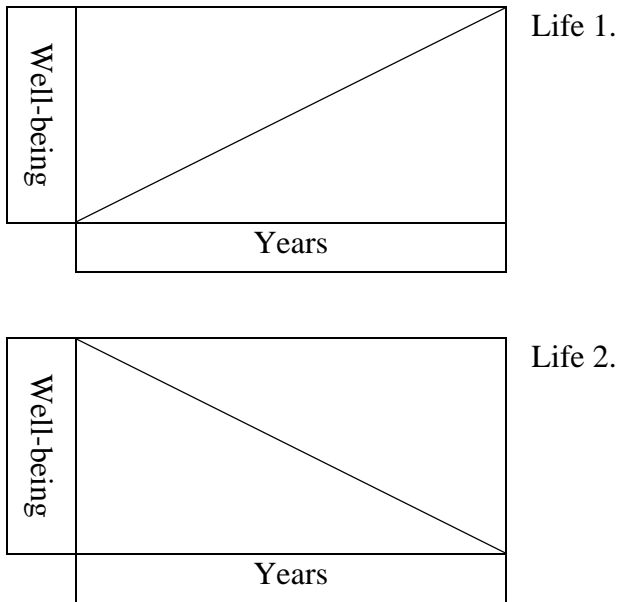
Slote, in chapter One of his *Goods and Virtues*, at least hints at making literary quality a component of the goodness of a life.²⁵ Slote claims: "Human life seems, as I have said, to possess a natural, though socially influenced, development of different times or stages of life . . . I believe that such a division [youth, adulthood, and old age] into 'times of life' tends to be accompanied, in most of us, by a sense of the greater

²⁵ Technically, the "goodness of a life" is not necessarily the same concept as "well-being." If Slote is after a different concept than the concept of well-being, then what he claims of "the goodness of a life" can be extended to "well-being" as a thought experiment. If Slote is after the same concept that I am, then of course what he says applied directly. Whether direct or indirect in application, Slote's arguments and ideas are worthy of critical examination, if for no other reason that to clarify my own position.

importance or significance of certain times of life in comparison with others” (Slote pg. 13-14). Slote calls his view, the view that the timing of a success (achievement, or good) matters to how good a life is, a “time preference” view. Slote does not necessarily commit himself to the claim that the good of a person must have a *literary* quality, but as I understand the time preference view, he comes close. Slote claims that successes in “the prime of life” are comparatively more important to those of childhood and old age (Ibid, pg. 18, 26). He also suggests that there are certain activities that are appropriate to each period of a life (Ibid, pg. 19-21).

David Velleman, in his “Well-Being and Time,” builds on Slote’s views discussed above. Velleman goes further than Slote explicitly to claim that the amount of well-being one has depends on the “narrative or dramatic relations” of events in one’s life (Velleman 1991, pg. 49). However, to describe Velleman’s view of well-being by only looking at the aspect that includes narrative or dramatic relations (or life stories) would be incomplete. For Velleman sees well-being as radically split into two distinct components: momentary well-being and well-being over time. On this distinction, Velleman says: “I therefore favor the principle that a person’s self-interest is radically divided, in the sense that he has an interest in features of his life that aren’t at all reducible to, and hence cannot be exchanged with, patterns of momentary well-being” (Ibid, pg. 61-62). Velleman calls momentary well-being a first-order good while well-being over time is a second order good. A second order good is a “valuable state of affairs consisting in some fact about other goods” (Ibid, pg. 58). Second order goods are irreducible to first order goods if they “at least possess value over and above that of [their] component first-order goods” (Ibid, pg. 58). The basic idea that Velleman is after

is that we can imagine two lives that have the same amount of overall momentary well-being, but different amounts of value (temporally extended well-being). A more detailed discussion of Velleman's views will come later. For example, here are graphs of two lives with the same amount of overall momentary well-being:



Although the momentary well-being of lives 1 and 2 differ at almost every single moment, the overall momentary well-being, when summed up at the end of both lives is the same. The basic intuition that Velleman is trying to get at is that the first life is more valuable than the second, even though they have the same amount of momentary well-being.

Velleman goes further than Slote because while Slote's explanation of the time preference view allows for a formula that adds (or multiplies) goodness at each stage of life thus resulting in increases at particular periods and not others (as I explained above), Velleman's does not admit of taking momentary well-being and applying a formula to generate well-being over time. As he says:

Some of the value judgments considered above are incompatible with any reduction of diachronic well-being [well-being over time] to synchronic well-being [momentary well-being], no matter how sophisticated an algorithm of discounting and weighting is applied. Because an event's contribution to the value of one's life depends on its narrative relation to other events, a life's value can never be computed by an algorithm applied to bare amounts of momentary well-being . . . (Ibid, pg. 60)²⁶

The narrative relations between events (which are required to get diachronic well-being) suggest a much more complicated model than suggested by the charts above. The charts above only deal with *amounts* of well-being at any given time in hypothetical agents' lives. Velleman requires a more substantial account of the *content* of the life in order to assess its diachronic well-being.

He does not provide a formula that would enable us to determine the amount of diachronic well-being in a life, given a detailed sequence of synchronic well-being. Such a general formula would have to involve a complex weighing of narrative and dramatic relations.

I certainly do not think that just any literary quality of a life would enhance a person's prudential well-being. Imagine that the Count's life ended in the prison; we would certainly call his life a tragedy. Tragedies have a literary quality of their own, but it would be absurd to say that a life full of tragedy is thereby enhanced with respect to prudential well-being. As was mentioned above, tragedies are beautiful, not ugly, so the aesthetic value of a tragic life would be positive. But what makes a tragedy a tragedy is that something goes horribly wrong in a life. In many tragedies, the tragic element is an extreme decrease in prudential well-being.

²⁶ Velleman calls his account of well-being over time "strongly irreducible" (Ibid, pg. 60). Slote's time preference view would allow for some level of reducibility: Velleman says that Slote's time preference view has "weak irreducibility" (Ibid, pg. 60).

What if we restrict the kind of literary quality that we take to enhance well-being in a way that excludes tragedy? Now imagine that the story ends with the Count finding the great treasure. This life would have a literary quality to it—but it would not be a tragedy. Let me say his life story in this case would be an “uplifting drama.” Is the life story in an uplifting drama a story of a life that thereby has an enhanced level of well-being? No, I do not think so.

A preliminary worry is that throughout the third proposal, no fuss has been made of the distinction between a life and a life story; only the second can properly be said to have a narrative structure. How could a life go well in virtue of having a life story of a certain sort? This seems a puzzle to which I have not heard an adequate account and do not have even a clear idea of how an account would go.

Secondly, the restriction to “uplifting dramas” appears *ad hoc*. If we were trying to distinguish between beautiful and ugly life stories, trying to find a distinguishing characteristic would not be *ad hoc*, for we would then be trying to find the distinction between positive and negative aesthetic values. However, the distinction between uplifting dramas and tragedies is of a different sort. Here, we are trying to distinguish between two sorts of beautiful life story. I can see no real motive for building this distinction into the account of aesthetic value apart from trying to fit it into prudential well-being. Both uplifting dramas and tragedies are aesthetically valuable on initial construal for the third account of aesthetic value. It is only after it is clear the account of aesthetic value could not plausibly serve as an account of what enhances one’s prudential well-being that the distinction between uplifting dramas and tragedies is made. There must be some sort of external, non-*ad hoc*, and principled reason for the distinction. The

proponent of the third account cannot claim that it is having a beautiful life story, as such, that increases prudential well-being. After all, both tragedies and uplifting dramas can make for beautiful life stories.

Thirdly, it is not clear that we could properly classify the literary types *tragedy* and *uplifting drama* without recourse to the concept of prudential well-being. So if we refer to the narrative relations of an uplifting drama in our account of well-being, our account may be circular. For example, the literary quality of an uplifting drama could require lows before highs, while a tragedy could require highs before lows. If these dramatic relations *require* that there be some variation in prudential well-being, some dip below a high level of well-being and an overcoming of some obstacle, then we must have some independent way of getting at the concept of prudential well-being.

Although it may seem as though I disagree with much of what Velleman says, I actually only disagree with the thought that prudential well-being could be constituted, wholly or in part, by the narrative and dramatic relations between events in one's life, where these narrative and dramatic relations are a matter of its aesthetic value. That leaves a lot of room for agreement. Velleman thinks that well-being is radically divided between diachronic and synchronic well-being and that diachronic well-being is constituted by the events that compose synchronic well-being plus the relations between these events. I can agree with this. So when Velleman says that well-being is radically divided, I can agree with him—in a way, but with a different distinction and for different reasons. I can agree with the idea that there is diachronic well-being, and even diachronic prudential well-being. I think, however, that the diachronic prudential well-being must be determined by something *other than* narrative and dramatic relations.

So none of the proposals for aesthetic value I have critically discussed turn out to be plausible candidates for something that would enhance prudential well-being by partly constituting it. There may be some further proposal about aesthetic value and its relation to prudential value. I feel however, that the burden of proof is on those who wish to claim that aesthetic value is a constituent of prudential well-being.

1.6 Conclusion

I would like to return to a distinction I made above between two different concepts of well-being and compare the theories of Velleman, Brink and Darwall. On the one hand we have a broader concept where something like the “good life” is meant by “well-being.” On the other hand, we have the narrower concept where something like “prudential value” is meant. If I understand Velleman correctly (and this is somewhat tentative, because he only has room to give a sketch of his overall theory), he is after a broader concept of well-being. Brink and Darwall use “well-being” in the broad sense to include at least prudential and moral goodness. Velleman may use “well-being” in a different way to include prudential and aesthetic goodness.²⁷ If Velleman, Brink, Darwall are after a different concept than I am, then in some ways, we might not really disagree about “well-being.” I can agree with what they say about the broader concept of well-being and they could agree with the narrower concept.²⁸

At this point, we should have a general idea of what non-moral, non-aesthetic, self-regarding interests are. I claim that satisfaction of such interests is constitutive of

²⁷ Darwall’s phrase “what we would rationally want for someone were we to care for them,” as I understand it, is a characterization of overarching value in a life. So if aesthetic value is indeed a separate and basic value, then it too would be a feature included in Darwall’s concept of well-being.

²⁸ Interestingly, even though they are after different concepts than I am, I will discuss the plausibility of their conceptions of well-being as candidates for the narrower concept I am after. So we will see the work of Darwall, Brink and Velleman again.

prudential well-being. In what follows in the next chapter, I will argue for a particular *conception* of the concept of prudential well-being. In the subsequent chapters, I will continue to develop and defend what I take to be the best conception of prudential well-being.

CHAPTER 2

CONCEPTIONS OF WELL-BEING

Having just distinguished between a few different concepts of well-being, and made clear that I am after the concept of prudential well-being, I now turn to the conceptions. First I will give relatively brief accounts of the various conceptions. Next, I will explore the various conceptions by applying them to the two primary protagonists in Mary Shelley's *Frankenstein*, among other hypothetical cases. Lastly, I will critically analyze the conceptions in light of what they imply about these cases.

I will adopt a four-part taxonomy of the basic conceptions of well-being. There are alternative taxonomies, but this one is the best for my purposes.¹ The four groups of conceptions of well-being are as follows: (1) mental state theories, (2) desire-satisfaction theories, (3) explanatory objective theories, and (4) pluralistic theories. Mental state theories treat some mental state, such as pleasure or happiness, as constitutive of well-being.² Desire-satisfaction theories treat the satisfaction of desires, or the satisfaction of

¹ Two alternative taxonomies are: (1) a two-way taxonomy as found in Griffin (1986) and Sumner (1996), and (2) a three-way taxonomy as found in Parfit (1986), Scanlon (1998) and others. The two-way taxonomy traditionally divides the conceptions into subjective and objective groups. But what features one uses to make the distinction is not an easy and obvious issue. The three-way taxonomy as found in Parfit and others leaves an important group of theories out. There is an alternative four-way taxonomy in Shelly Kagan (1992). Kagan's four-way taxonomy is composed of two distinctions which cut across each other (subjective/objective and internal/external) making for four basic theories. This four-part taxonomy may make things a bit clearer, but the *importance* of the categorization is unclear. Additionally, Kagan's taxonomy still leaves out an important group of theories (what I call "pluralistic" theories). Now, I do not think that the other two other types of taxonomy are wrong, strictly speaking, but I do think them lacking in the virtues of my four-way taxonomy.

² I do not in this dissertation use "constitute" as some philosophers do. Some philosophers think of "x constitutes y" as asserting a specific metaphysical relation between *x* and *y*—a reductive metaphysical relation. In this dissertation I use "constitute" in a way that is compatible with reductionism, yet does not commit me to reductionism. I attempt to capture what some philosophers call the "making" relation by my use of "constitutes."

some relevantly construed desire set, as constitutive of well-being. Explanatory objectivists treat a *single* feature, such as a person's developing or exercising her human nature, as constituting well-being.³ The first three families of conceptions of well-being are *monistic* in that they treat some single feature as constituting well-being. The fourth type of conception of well-being is *pluralistic* in that it treats multiple factors as constituting one's well-being. Pluralists, on this characterization, treat each of the factors as independently constitutive of well-being. In a later chapter, I will discuss what I call "mixed theories" of well-being that combine features from theories in more than one group. For mixed theories, well-being only increases when two or more features combine. Mixed theories qualify as monistic because although the feature that constitutes well-being according to them is not *elemental*, they are still best thought of as picking out only a single feature that constitutes well-being, though that feature is *complex*.⁴ Pluralists treat each factor as self-sufficient for constituting well-being.

2.1 Four Conceptions

2.1.1. Mental State Conceptions

Let us first examine mental state conceptions of well-being. A person's well-being, on mental state accounts, is composed exclusively of some set of mental states of the person. Accordingly, a person's well-being at a time is set by his or her occurrent mental

³ I draw the terms "explanatory objectivism" from Kitcher's "Essence and Perfection." Please see 2.1.3 for a more detailed account of this sort of theories.

⁴ For what it is worth, I confess that the distinction between "complex" and "elemental" features is fuzzy. I have set up the *arguments* for and against the conceptions so that little, if anything, depends on the distinction. The only thing of this dissertation that depends on the distinction is the *organization* of the chapters: I deal with elemental monistic theories in this chapter and complex monistic theories in a later chapter. So the distinction really just ends up affecting when I deal with the theories rather than how I deal with them. Also, as a bit of an aside, one should recognize that it is possible to develop a pluralistic conception of well-being composed of mixed features.

states at that time. Thus, the presence of the correct kind of mental state in the past is not constitutive of well-being in the present. Naturally, there are restrictions as to which kind of mental state matters for well-being. Two very popular candidates are pleasure and happiness. Of course, “pleasure” and “happiness” are themselves, at least to some extent, up for definition as well. Philosophers have taken the terms to mean many different things. I do not have time to pursue the history of the conceptions of these two mental states, if indeed they are even different. But some further exposition is necessary.

The strategy that virtually all mental state proponents adopt is to assert that well-being is pleasure or happiness and then goes on to characterize pleasure or happiness.⁵ Sometimes pleasure is thought to be a sensation one has (hereafter, the sensationalist view). The idea that happiness is a sensation is a bit less plausible than the idea that pleasure is a sensation, but if one thinks that pleasure and happiness are the same and that pleasure is a sensation, then one is committed to thinking that happiness is a sensation. Upon tasting a cold beer on a hot day, one may experience a pleasurable feeling. Another example is the pleasure one feels when listening to a great piece of music. Apart from the sensationalist view, there is another view: pleasure and happiness are attitudes (the attitudinal view). More specifically, the view is that pleasure and happiness are propositional attitudes in that they take propositions as their objects. So, for example,

⁵ Another strategy is available. One could “cut out the middle-man” and just claim that well-being is composed of certain kinds of mental state and then just go on to characterize this mental state—all of this ignoring whether or not such a state is pleasure or happiness. I have not come across anyone who adopts this strategy for a monistic theory—everyone who adopts a monistic theory seems to like to keep pleasure or happiness in the middle. There are some pluralists, as we will later see, who pick out other mental states as relevant to well-being, such as aesthetic appreciation; but these other mental states are on the list of things that make a life go well *in addition to* pleasure or happiness.

one can be pleased that one's daughter is doing well at college or one can be happy that the Royals baseball team are in the pennant race.⁶

Happiness, when conceived to be a propositional attitude, might be taken to be a factive mental state. What I mean if I were to say "happiness is factive" can be illustrated with an example: in order for a person to be happy that his daughter is doing well in college, his daughter must really be doing well in college. So, if happiness is factive in this way, one can be happy *that* such and such is the case only if it really is the case. If his daughter is not doing well in college, the father might be construed as happy *in the thought that* his daughter is doing well at college. Being happy "in the thought that" such and such is the case allows for that thought to be mistaken or wrong. If happiness is factive, however, being happy "that" such and such is the case makes no such allowance. If happiness is factive in this way, then the important thing to keep in mind is that the mental state is what really matters here. And the mental state is relevantly the *same* between (1) happiness that *x* is the case and (2) happiness in the thought that *x* is the case.⁷ The only thing that is different between (1) and (2) here is the way the world is. One *could* deny that happiness is factive in this way. But I do not want to get into the debate here. All that I am claiming is that whether we treat happiness or pleasure as

⁶ So, what is it that makes both the sensation from the beer and the sensation from the music both fall into the same category of pleasure? Presumably it is not because they are the *same* feeling. They are the same type of feeling; that is to say, that they fall into a type called pleasure. But then what makes the category of pleasure? The same can be asked of the attitudinal view as well—what makes all of those attitudes pleasure or happiness?

⁷ Actually, if happiness is factive, those who are happy that *x* is the case (on this factive interpretation, *x* must really be the case), strictly speaking, are also happy in the thought that *x* is the case. I qualify this by "strictly speaking" because when one says that someone is happy in the thought that *x* is the case, it might be conversationally implied that *x* is not the case; but, of course, implication and literal meaning are two different things. So on the factive view, it could be argued that what is important is *happiness in the thought that x is the case*, whether *x* is actual or not. In any case, the result is the same, that the state of mind is what is important.

factive mental states or not, the important thing is the mental state as such. So, even if happiness is factive, happiness that x is the case and happiness in the thought that x is the case are both treated as increasing well-being.⁸

Something to keep in mind regarding mental state accounts is that they do not necessarily have to play down to the lowest and basest human inclinations. Often, critics of mental state accounts, especially when attacking views that associate well-being with pleasure, chastise the views for pandering to the lowest common element in people; for example, sheer sensory pleasures one gets from food, sex, and even drugs.⁹ But the relevant class of mental states can be construed as far more inclusive than this objection implies. Of course, as the class of relevant mental states expands, getting right the account of what makes all of the disparate feelings and attitudes pleasures becomes increasingly important.¹⁰

The key attractions of the mental state accounts are, I think, two-fold. The first is that the presence of the right kind of mental state plays a crucial role in how, pre-philosophically, we evaluate our own lives. Very often it is the case that, in an ordinary conversational context, when a person is asked how well so-and-so is doing, his reply involves some comment on how so-and-so is feeling. The second attraction of the mental state accounts is that, according to them, each person has a very good idea of how well

⁸ Later in this dissertation, though not in this chapter, I will study a theory that treats only factive happiness as increasing well-being. However, such a theory is no longer a pure mental-state theory, as it includes a reference to something outside the mental realm. I must put the “mixed” theory aside for now.

⁹ This objection to the hedonistic accounts of well-being could be characterized as the “fit for swine” objection.

¹⁰ Of course, one could say that nothing holds them together as pleasure or happiness. This would be to take the strategy as found in the fourth footnote of this chapter.

off he or she is.¹¹ This tight connection between a person's beliefs and his or her well-being is intuitively attractive because well-being, seems, in a strong sense, to be in and of the person's lived experiences—well-being and a person's experiences seem closely linked.

2.1.2 Desire Satisfaction Views

The desire satisfaction account of well-being is—to a first approximation—that one's level of well-being is set by the extent to which one's desires are satisfied. A desire is satisfied if the desired state of affairs obtains. For example, if my desire is that the Royals win the pennant, then my desire is satisfied if the Royals win the pennant. It is worthwhile to note the possibility that a desire can be satisfied and yet the agent not know, or even believe it. For example, I can desire that the Royals win today's baseball game and it can be true that the Royals have won today's game. In this case, the desire might be satisfied without my knowing or believing it.¹²

Just as the mental state theories take great pains to characterize which kind of happiness is important for well-being, so too does the desire-satisfaction proponent take great pains to characterize which desires matter. Desires are often characterized quite broadly here to include anything from whims one might have to life-long plans and projects. Sometimes all of these states are referred to as pro-attitudes. So a desire, on

¹¹ Now, sometimes one hears of people saying "I thought I was happy, but I wasn't" or "I didn't know how happy I was." If indeed these are real cases of the person being wrong about how happy he or she was, then the connection between the person's mental awareness and her happiness is less than perfect. I tend to think that they rarely should be taken literally, though there do seem to be at least some instances when they hold true. However, the connection between a person's judgments about her well-being and her actual well-being still seems strong in most cases and that is something that, at least intuitively, favors the mental state accounts.

¹² Additionally, one can come to desire something that is already the case. For example, someone, call her "A," can come to desire that her friend have a romantic interest in her when A's friend already has such an interest. In this case, A's desire is satisfied immediately.

this view, is not just a felt tug or pull at a given moment—though of course, tugs and pulls are still desires. It is important to be charitable to those who endorse mental state accounts by not construing the relevant mental states as just base sensory pleasures. Similarly, it is important to understand desires to be far broader than just the urgings that someone feels at any given moment. So, on this broader view of desires, one could have a desire for x without, at that very moment, having some urge to go out and get x .

The desire-satisfaction theory is different from the mental state theory in one very important way. Whereas the mental state theory takes a person's well-being at a time to be composed entirely of occurrent mental states, the desire theory only treats desires as necessary conditions. Desire-satisfaction theories then look to the world to see whether the desired states of affairs obtain or not. If the desired states of affairs do obtain, then the agent's well-being is enhanced.¹³

As we will see in the next chapter, some desire-satisfaction accounts give a counterfactual analysis of well-being in terms of what an agent would desire under certain conditions, so they do not even treat actual mental states as necessary conditions of well-being. This means that there is a family of desire-satisfaction accounts rather than merely one. Yet I will simplify my discussion for now by talking as if there is only the one simple view. The more complicated views are better left for a later time.

The attractions of the desire-theory are, I think, two-fold. First, like mental state accounts, desire and desire satisfactions play a key role, pre-philosophically, in how we see and judge our own lives. For example, countless stories in the literary and religious traditions count desire-satisfaction as one of the central aspects of a life's going well—

¹³ I will discuss a slightly different account of desire-satisfaction in Chapter 7 of this dissertation. For now, I merely want give the reader an idea of in what desire-satisfaction consists.

(e.g., Job, who loses his wealth and children and is struck with boils, is someone who loses what he desires most and is often thought to have his life go prudentially worse thereby). There is also the tale of the Jinii who grants wishes in such a way that it leads to the thwarting of other wishes and thereby make “the wisher’s” life go worse. Both mental state theories and desire-satisfaction theories seem to match up very nicely with our pre-philosophical intuitions about how well lives go.

Now I will move on to what I take to be the second key advantage of the desire-satisfaction theory. Please recall that desire satisfaction theories require no belief, on the part of the agent, that the desire is satisfied or that object proposition is true. That is, for the satisfaction of a desire that *p* to enhance well-being it is not required that the agent believe that *p*. Moreover, according to the desire-satisfaction theory, one’s well-being could diverge from one’s thoughts about one’s well-being, just because one could be wrong about whether one’s desires are satisfied. For example, someone could *desire* that her friend be a sincere and genuine friend and *believe* that her friend is sincere and genuine and yet be *mistaken* in that belief. On the “simple” desire-satisfaction account, her well-being is lower in virtue of that thwarted desire.¹⁴ With the desire-satisfaction account, there is a tight connection between a person’s well-being and aspects of her psychology. But the desire theory does not go so far as to make well-being constituted wholly by a set of mental states; there must be some connection to the world. Given the above example, there seems to be a great advantage in that connection to the world.

¹⁴ Quite probably, there are other closely-related desires that are thwarted as well; such as the desire not to be lied to, etc. So, the agent’s well-being could be lower still in virtue of some other desires not being satisfied.

2.1.3 Explanatory Objective Theories

Explanatory objective theories have a long history of supporters in philosophy, and have had a relatively recent renaissance in the literature. They typically base well-being in human nature. Given its recent revitalization, there are not many recent full-blown expositions of this view.¹⁵ In light of this, my strategy in 2.1.3 will differ from those in 2.1.1 and 2.1.2. Whereas in 2.1.1 and 2.1.2 I rarely mentioned any particular author, in the present part, I will do so liberally.

Before getting into the details of objective accounts and human nature, I would like to remind the reader of my preferred terminology and the theories associated with each term. Philip Kitcher in his paper “Essence and Perfection” distinguishes between bare objectivism and explanatory objectivism. Explanatory objectivism unifies the various things that make one’s life go well by citing a fundamental element that figures centrally in the account of what constitutes well-being. Human nature plays this role for most objectivists of this sort. Bare objectivism, in contrast, has no single unifying element; each element contributes to well-being. So bare objectivism is non-explanatory objectivism. Bare objective theories are sometimes called “objective list theories.”¹⁶ What Kitcher calls “bare objectivist” theories and what others call “objective list” theories, I prefer to call “pluralistic” theories. I prefer to stick with the descriptor “pluralistic,” because it captures something more about the set of theories under discussion than does “objective list” and “bare objective.” When the word “objective” is used to identify a class of conceptions of well-being, one might confuse it with

¹⁵ Hurka’s book *Perfectionism* is one recent notable exception.

¹⁶ “Objective List” as a name for the theories under discussion was popularized by Parfit in his “What Makes a Life Go Best” in his *Reasons and Persons*. Several philosophers have followed Parfit in his choice of terminology. I find the name “pluralism” to be preferable.

explanatory objectivist accounts. The motivation for explanatory objectivism and pluralism differ so much that grouping them together would no doubt rely on some arbitrarily chosen similarity between the groups, instead of identifying and distinguishing the groups based on core or central characterizations.

Most explanatory objective theories claim that increases in well-being come from developing and acting in accordance with human nature. There could be other sorts of explanatory theories, but alternatives to human nature theories are rarely defended.¹⁷ Theories of human nature are, of course, numerous. Some have thought that human nature is that which is distinctively human, but it seems that distinctiveness is not all that important. For if distinctiveness is understood as biological, then it seems relatively unimportant (especially for a theory of well-being). Take for example, the distinctive aspect of humans as featherless bi-peds. It is hard to see how this could play a role in an account of well-being. Alternatively, if distinctiveness is taken to be some higher level human functioning that is intuitively much more relevant to well-being, then it is not clear why other possible creatures could not have that level of functioning as well. For example, take rationality. Being rational seems much more relevant to well-being than does being a featherless biped, but it could be that there are other creatures who are rational.

Rather than focus on distinctiveness, a better focus is on the human essence if there is such a thing. The human essence, on one proposal, consists of a bodily essence and

¹⁷ Thomas Carson in his *Value and the Good Life* provides a contemporary defense of a divine preference theory, according to which one's life goes well to the extent to which one lives in accordance with what God wants for one (pg. 239-254). I treat Carson's theory as a sort of desire-satisfaction account, though of an unusual pedigree, instead of an alternative sort of explanatory objective account.

rational activity, both practical and theoretical rational activity.¹⁸ Bodily essence could be described as that which humans share with other animals. To develop it, one needs food and exercise. The two kinds of rationality are perhaps not completely unique to humans. But certainly the level of rationality of humans is almost incomparably greater than that of the other known creatures. Practical rationality is reasoning about what to do. Theoretical rationality is reasoning about what to believe. The view of this kind of objectivism is that one's life goes well to the extent to which one exercises and develops these elements. According to the theory under discussion, development and exercise of one's bodily and rational capacities only counts as increasing well-being because they are the components of human nature. If human nature were to be different, then one's well-being would not be enhanced by the development and exercise of these capacities. Thus, although the theory may appear pluralistic in that it picks out two features as relevant for well-being, it really is monistic because those features are relevant to well-being *only because* they fit into an account of human nature.

This is all, of course, controversial. The temptation to give an evolutionary or socio-biological account of human nature is very strong here. There is a great deal of *prima facie* intuitive plausibility to an evolutionary account of human nature because after all, an evolutionary account would involve a story of how humans came to be as they now are. There are serious problems with any theory that treats well-being as consisting in the development and exercise of one's nature when one understand human nature as given by an evolutionary or socio-biological account. I will discuss this issue in 2.2.2.

¹⁸ This is drawn from Hurka's *Perfectionism*: to my mind, the most plausible and detailed of the objective accounts.

Interestingly, some recent proponents of human nature theories have suggested that there is room for substantial individual difference in what makes a life go well.¹⁹ There are very deep questions about whether human nature is set as one nature for all members of the group of humans or whether there is room, with the correct account of human nature, for there to be substantial individual difference. There might be a single human nature, on some accounts, but a theory of well-being might still allow for variation. The risk, of course, is that in the accounts that allow for substantial differences in what makes people's lives go well, the accounts might not be of human nature any more, but of some other feature. I must leave critical discussion of this topic for 2.2.2.

The intuitive plausibility of explanatory objective accounts is two-fold. Well-being, on most explanatory objective accounts, consists in living life in accordance with, and developing, one's nature. The first advantage is that the accounts make reference to the essence of the person. What could be a better account of one's well-being, so the thought goes, than whether one is living his or her life developing and exercising what is at the very center of the person? This connection to the essence of the person does not have the pre-theoretical and pre-philosophical intuitive pull of the mental-state and desire-satisfaction theories, because so much groundwork must be laid to get the idea of a human essence even on the table. However, with a plausible account of human nature, the connection to one's nature is attractive. The second intuitive advantage of explanatory objective theories is that they have a characteristic that is very similar to

¹⁹ Rasmussen, in his "Human Flourishing and the Appeal to Human Nature" is an excellent example of someone who is what I call a explanatory objectivist and yet thinks that there is a great deal of room for individual differences in human nature. As the title of his article indicates, Rasmussen is after the concept of human flourishing—which seems a different and broader concept than the concept of prudential well-being. I will treat Rasmussen's work as if it were about the narrower concept however, as I do with several of the so-called explanatory objectivists.

desire-satisfaction theories; they both tie well-being to the world in which the person lives. One's well-being, according to human nature and desire-satisfaction theories, depends on the state of the person and the state of the real world. Both theories make well-being not depend entirely on mental states, as do the mental state theories. The mental state theories have attractions of their own, but a connection to the outside world is surely something that helps make human nature theories initially plausible.

2.1.4 Pluralistic Theories of Well-Being

Pluralistic theories are quite different from the monistic accounts discussed above. Pluralistic theories have a very different structure and motivation. I will discuss two notable recent accounts of pluralistic conceptions of well-being.

James Griffin in his *Value Judgement*, offers a list of five items that make a life go well. Griffin sets up his list with the following comment: "To see anything as prudentially valuable, then, we must see it as an instance of something generally intelligible as valuable and, furthermore, as valuable for any (normal) human. Prudential deliberation ends up, I think, with a list of values" (Griffin 1996, pg. 29). Griffin's account of prudential value makes it necessary for the valued thing to be intelligible as valuable—"intelligible to whom?" is of course a big question.²⁰ Also, what role "normal" and "human" play is also quite open ended. But what should be clear is that Griffin is generating a list with normal humans in mind.

²⁰ Griffin's comment makes me think of a character from Jane Austen's *Emma*, Mr. Woodhouse (Emma's father). Mr. Woodhouse complains that his son-in-law is too rough with his children—always tossing them up in the air. Emma replies "But they like it, papa; there is nothing they like so much. It is such enjoyment to them, that if their uncle did not lay down the rule of their taking turns, whichever began would never give way to the other." Mr. Woodhouse says "Well, I cannot understand it" to which Emma replies "That is the case with us all, papa. One half of the world cannot understand the pleasures of the other" (Austen, *Emma*, ch 9, pg. 624).

Griffin's list has five items: (a) accomplishment, (b) the components of human existence (autonomy, capability, and liberty), (c) understanding, (d) enjoyment and, (e) deep personal relationships. Griffin thinks of his list as being a something like a rough draft of the list; with possible amendments to come later (Ibid, pg. 30).

John Finnis, in his *Natural Law and Natural Rights*, offers a list of things as an account of "'good,' 'basic good,' 'value,' 'well-being'" (Finnis, pg. 86). Finnis' list is: (1) life, (2) knowledge, (3) play, (4) aesthetic experience, (5) sociability (friendship), (6) practical reasonableness, (7) religion. Finnis characterizes the items on the list very broadly. For example, Finnis describes knowledge as: "the activity of trying to find out, to understand, and to judge matters correctly" (Ibid, pg. 60). Finnis is far more certain about his list than is Griffin, Finnis thinks of his list as something like a final draft. Finnis claim that each element is non-reducible to any other and that the list is exhaustive of well-being. Finnis takes himself to be providing a list of items for *human* well-being (Ibid, pg. 92-95).

Notice that Griffin's and Finnis' theories appear to be pluralistic theories. One thing that Griffin and Finnis are *not* clear on, frustratingly, is whether they are full-bodied pluralists. Recall that what is meant by "pluralism" here is that each element on the list is, by itself, sufficient for at least some well-being. If they are pluralists in this sense then a person could have some amount or degree of just one of the items on the list and thereby have some amount or degree of well-being. Perhaps, though, Finnis and Griffin are not pluralists in this full-bodied sense. Perhaps one needs to have at least a little of each item on the list to have any amount of well-being. On this construal, some degree of

each item is a necessary, rather than a sufficient, condition for having some degree of well-being. I cannot resolve these problems here. I will treat the authors as pluralists.²¹

The motivation for accepting pluralistic theories of well-being is that they are so very inclusive. A pluralist of the sort I have in mind has the strategy of “having one’s cake and eating it too.” If the mental state, desire-satisfaction, and explanatory objective theories each leave something out (in different ways), then a pluralist can include aspects from each theory and with luck, do away with any shortcomings.

2.1.5 Conclusion

Each of the types of conceptions has what we might think of as an attractive starting point. There is at least an air of initial plausibility with each. That is not to say that one is just as good as the next, but there is something to each of them. What I will do in 2.2 is to take a deeper look at the advantages and disadvantages of each of the conceptions. Once the skin is peeled away, desire-satisfaction theories will turn out to be best.

²¹ Might Griffin and Finnis really be explanatory objectivists disguised as pluralists? This is an intriguing question. There is some chance that they are. Presuppositions about human nature and the world in which humans live play a role in each philosopher’s thinking. Griffin, in a very helpful footnote on this issue, writes:

My own list is very much out of a particular tradition: modern, Western, and atheist. But take someone with a radically different list: instead of enjoyment, the mortification of the flesh; instead of deep personal relationships, cloistered solitude; instead of autonomy, submission to the will of God. But of course one’s list will change with one’s metaphysical views. If I believed in a certain kind of God, I too might have a different list. But this makes lists relative not to culture, but (unsurprisingly) to one’s judgements about the world one thinks one is living in. (Griffin, pg. 15)

Griffin’s footnote is quite frank in expressing the difficulties in coming up with a list of items. Perhaps Griffin does presuppose an account of human nature when coming up with his list. Finnis explicitly states that his list is only for humans (Finnis, pg. 92-95). Griffin, in a quotation above, makes a similar claim. So there is some reason to believe that they are really explanatory objectivists at some deep level. However, at this point, I will treat Griffin and Finnis as full-bodied pluralists.

2.2 Critical Discussion of Conceptions

2.1.1 The Story

Monte Cristo's life is not a great case study for critically assessing different conceptions of well-being. This is because the four different conceptions diverge only slightly in what they would say about Monte Cristo's well-being. Given the purpose of Chapter 1 (to isolate the narrow concept of prudential well-being), the example was suitable. Now, however, I need cases where the different conceptions come apart. Things will seem more controversial, and one's intuitions should not be so definite here. To fit my needs, I have chosen one of the greatest stories of internal and external conflict in literature.

The case study I have in mind is of the two primary protagonists in Mary Shelley's *Frankenstein*. The first thing I would like to do is to dispel the common thoughts about the book and characters. If the reader is not familiar with the book, forget all that you think you know about Frankenstein. Victor Frankenstein, the scientist, is the creator of his "monster," who is never given a name. Frankenstein, a genius chemist and physiologist, discovers a method of bringing dead human matter back to life. He gathers parts from different corpses and then composes something of a super-human. He brings his "creation" to life, though without the aid of lightning. The monster is extremely strong and agile and is very intelligent. He is superior to normal humans in all of these respects. However, the monster is also repulsively ugly. Frankenstein, upon bringing life to his monster, panics and abandons the creature.

Frankenstein's monster is left to fend for himself. He manages not only to survive, but also to become a master of language and learn a great deal of history and some of the other liberal arts. All of this learning is accomplished without direct assistance from

humans, although he “eavesdrops” on a family that lives in a cottage. When he tries to confront people, they are repulsed by his ugliness and he is driven away. The monster, created with a very strong longing for companionship and a great sensitivity to rejection, eventually tries to find his creator, Frankenstein.

The purpose of the monster’s search is to convince Frankenstein to create another being—another monster, but this time a female monster who is also ugly and super-human in many ways. The monster finds Frankenstein and convinces him to make another being.²² Frankenstein, out of fear of horrible consequences, and at the last minute, tears up the still life-less body of the monster’s would-be companion. The monster then becomes enraged (again), and kills several members of Frankenstein’s family including Frankenstein’s bride on their wedding night. Frankenstein dedicates the rest of his life to destroying his monster but dies in the effort. The monster, whose sole aim towards the end of his life was to torment Frankenstein, then goes off to die alone.

The idea I have here is that if we take the first three conceptions of well-being (mental state, desire-satisfaction, and explanatory objectivists) and apply them to the two life-stories, we will have different accounts of how well the lives are going. We can compare the accounts and try to decide which is the more plausible. I will criticize pluralistic accounts as well, though I think that pluralisms suffer from a general theoretical defect and so I will not argue against them using our intuitions based on the stories from *Frankenstein*.

²² This is slightly oversimplified, after the monster is rejected by everyone, he kills one of Frankenstein’s family members and frames another family member for the murder. Then the monster finds Frankenstein and convinces him to try to create a companion for him.

2.2.2 The Application of Explanatory Objectivism to the Story

There is a significant section of *Frankenstein* dedicated to the monster's acquisition of language and the learning of the liberal arts (chapters 12-15). It is in those chapters that the monster tells his tale of how he learned most of what he will ever know. He is able to find a hovel beside a cottage where a family of farmers lives. The family consists of a father (de Lacey) who is blind, and two children, a son (Felix) and daughter (Agatha). The monster is able to learn language by secretly observing the behavior of the family. He teaches himself to read and then goes on to read *Plutarch's Lives*, *Paradise Lost*, *Sorrows of Werther*, among a few other things. After this period of learning, he is extremely intelligent and is able to give lucid explanations of his history and very persuasive arguments to get what he wants.²³

Let us look first at the explanatory objective accounts that ground well-being in human nature. Such theories typically maintain that one's life goes well with the extent to which one exercises or develops one's human nature. In order to apply such theories of well-being, we have got to look at the nature of the Frankenstein monster. The monster is unusual, but any adequate account of well-being must be able to say something about him. For surely, his life can go better or worse. I think that the case of Frankenstein's monster will be a good test case precisely because it is such an unusual case. One might object that a dog's life can go better or worse as well and yet examining a dog's well-being will do little to assist our understanding of human well-being. I think that is true, but Frankenstein's monster and relevantly similar beings have lives that go

²³ He is so persuasive that after he has killed one of Frankenstein's relatives and framed another relative for the murder, he is still able to persuade Frankenstein to do a favor for him by making him a bride. The monster gives something of an explicit threat that if Frankenstein does not create a bride for the monster, he will harm him and his family. But this threat is accompanied by the argument that Frankenstein has an obligation to his creation to make him happy and that having a wife will pacify his great anger.

well or poorly in that they, too, have lives that can go better or worse with respect to prudential well-being. After all, they are rational and have many of the same mental states that human beings have.

A good place to begin is with an examination of the monster's nature. Let me return to an issue that was raised in the discussion of human nature earlier in this chapter. One's nature is given by one's essence, it seems. What is one's essence? There is a strong temptation to appeal to an evolutionary biological account, because such an account would involve a story about how we came to be the particular way we are.²⁴ There is a very serious problem, however, for any account of well-being that makes essential reference to human nature when human nature is understood in terms of evolutionary biology. There very well could be an evolutionary disposition to sacrifice oneself for one's genetic offspring. At first glance the thought may seem perplexing. However, especially in cases where there are sufficient numbers of offspring who could reproduce, or at least a few offspring who have a high chance of reproduction, and in a case where the parent could no longer reproduce, the greater chance of traits, characteristics, or genes continuing into the next generation is for the parent to sacrifice herself for her offspring. There are, of course, some species which reproduce so easily that dispositions of the parent to sacrifice for her offspring would not be selected in a competitive environment. However, humans do not reproduce at a high rate and must put a great deal of resources into each child. So humans could be the sorts of creatures that have a disposition for parental sacrifice. On the evolutionary account of human nature

²⁴ Indeed, the story is even more interesting for Frankenstein's monster because his story is a combination of evolution and scientific manipulation. But I leave that aside for the moment, in order to get to the nerve of a particular criticism.

and on the human nature account of well-being, such a sacrifice would not really be a sacrifice of well-being, but would merely seem so. The self-sacrifice would be merely illusory, according to the account under discussion, because the person would be living in accordance with his nature and would increase his well-being thereby. But, this is implausible. If one kills oneself for one's offspring, there normally would be a net decrease in one's well-being from the previous level of well-being. So, if one appeals to an evolutionary account of human nature, then human nature could not plausibly be viewed as playing a central role in what makes for well-being.²⁵ It is possible for one to act in accordance with his nature, on this account of human nature, and decrease his or her well-being, so acting in accordance with one's nature cannot always lead to increases in well-being. So the evolutionary biological account of human nature must be rejected by the most common sort of explanatory objectivists; though many people find such an account of human nature compelling.

Let us set aside these worries and, for the sake of argument, explore Thomas Hurka's account of human nature. Hurka thinks that one's life goes well to the extent that one exercises and develops one's human nature.²⁶ The best account of human nature treats human essence as central, according to Hurka. Human essence, on Hurka's view, has two components: physical essence and rational essence. Let us start by trying to find out the monster's physical essence. There are many difficulties here. The monster is a one-of-a-kind being, not a member of a type; more accurately, he is a single member of

²⁵ For a more detailed discussion of these issues see Kitcher's "Essence and Perfection."

²⁶ Strictly speaking, Hurka claims that he is not offering an account of welfare or well-being: according to him, he is offering an account of the human good. For what it is worth, I think that Hurka relies on too narrow a construal of welfare or well-being—he thinks that it is essentially subjective in some way. Regardless of whether Hurka relies on too narrow a construal of well-being, I adopt what he says to my discussion of well-being in this dissertation.

his own type. At least in many aspects, he is *sui generis*.²⁷ The fact that the monster is one of a kind does not mean he lacks a well-being. Indeed, the monster biologically human—though a composite of sorts. An account of human nature, such as Hurka's, must be able to explain the monster's physical essence. But what is it? Hurka's account of physical essence would seem to apply to the monster. Hurka thinks that minimally, the physical essence is good functioning of the organs and systems (Hurka, pg. 37).

Hurka remarks:

Each system in our body has a characteristic activity. The respiratory system extracts oxygen from air, circulatory system distributes nutrients, and so on. For a human to remain alive, each system must perform its activity to some minimal degree; for her to achieve reasonable physical perfection, it must do so to a reasonable degree. But a system does this when it is free from outside interference and operating healthily. So the basic level of physical perfection is good bodily health, when all our bodily systems function in an efficient, unrestricted way. (Ibid, pg. 38).

I do have some minor concerns about Hurka's account of physical perfection thus far. Presumably, medications that help the systems perform their characteristic activity or even improve the characteristic activity, such as medications that reduce blood pressure, increase physical perfection. And yet such medications surely constitute outside interference and these medications do restrict the systems in a significant way. So I not entirely satisfied with the way in which Hurka explains the basic level of physical perfection. But let me set aside concerns of this kind.

The account of physical perfection I have given so far is incomplete. There remains yet a higher form of physical perfection. Hurka thinks that great athletic feats count for more intrinsic value in a life. He says:

²⁷ Although he is extremely ugly, concerns of beauty are irrelevant here—they may play a role in the broader concept of well-being that I do not discuss in detail in this dissertation, but that is another issue. Aesthetics aside, he is far superior physically to anyone in existence.

Higher physical perfection comes in vigorous bodily activity. Here our major physical systems perform to higher degrees, processing more air, carrying more nutrients, and moving greater weights longer distances. This activity occurs most notably in athletics, and the Aristotelian perfectionism finds the highest physical good in great athletic feats. (Ibid, pg. 39)

Hurka admits that great physical feats also often embody practical rationality, but leave that aside for now. Hurka's examples of great athletic feats are the record-setting instances of the 100 meter dash and the long-jump. I am not entirely satisfied with how Hurka explains the highest kind of physical perfection. Some of his examples seem too broad; for example, it is not clear that in the case where one can process more air that one's well-being is thereby enhanced. Some of his examples seem too socially influenced; it is not clear why Hurka did not pick, say, the record-setting for hotdog eating instead of the record-setting long jump. I do not see any clear way to pick which of one's physical abilities it is in one's self-interest to develop apart from appealing to social norms, but this would mean appealing to something other than human nature.

To pick an example from *Frankenstein* to illustrate my problem with Hurka's theory, the monster speaks of having an excellent digestive tract. Perhaps that physical attribute should be developed and exercised on the perfectionist theory. I can see why developing and exercising his digestive tract is *instrumentally* good for the monster; then he could eat just about anything. But how could developing it be *intrinsically* good? The worry is that Hurka's account must pick out the features of one's nature to be developed, and this "picking" will either be arbitrary or we will be loading the questions about one's nature by relying on social norms. This is not meant to be a conclusive refutation of Hurka's account, but it is a worry that I think requires an answer.

Now let us move on to the non-physical aspects of the monster's nature. Does the monster have theoretical and practical rationality? Yes, and what is more, he is more

intelligent than normal humans. This is not the Frankenstein monster we grew up with in movies and television. He is more sensitive and quicker to learn than the vast majority of people.

An objective account, such as Hurka's, would measure his well-being to be extremely high throughout much of the novel. This is because the monster learns at an incredibly fast rate and he reasons extraordinarily well. Hurka's account implies that a person's well-being will increase with well exercised theoretical and practical rationality.

In Hurka's own words:

For theoretical perfection I will take this category [in which perfections occur] to be that of *beliefs*. A person's theoretical good at a time will depend on the number and quality of (some of) the propositions she believes at that time, so the issue is whether all her beliefs count or only those that are, say justified or true. On the practical side, the general category will be *intentions*. The relevant facts will concern the ends a person intends at each time or has resolved actively to pursue. (Hurka, pg. 101)

Hurka finds the number and quality of one's beliefs to be relevant to one's theoretical perfection. He finds the number and quality of one's intentions to be relevant to one's practical perfection.

Hurka is disinclined to treat all beliefs as enhancing one's perfection. He identifies four categories of belief an agent can have: (1) belief, (2) justified belief, (3) true belief, and (4) justified true belief (Ibid, pg. 103). These are, of course, not meant to be mutually exclusive categories; a mental state that meets (1) could also meet (4), for example.

Hurka has an analogous categorization for intentions. He says that there are (1) intentions in the belief that one will achieve the end, (2) intentions in the justified belief that one will achieve the end, (3) intentions in the true belief that one will achieve the end and (4) intentions in the justified true belief one will achieve the end (Ibid, pg. 103).

Hurka, for both beliefs and intentions, calls members of set (1) “attempts,” members of set (2) “deserving attempts,” members of set (3) “successful attempts,” and members of set (4) “deserving successful attempts.”

As mentioned above, Hurka does not hold that all beliefs and intentions count towards one’s perfection equally. He mentions two possible accounts of how much perfection one could attain in having a belief or intention in set (1) through (4). The first account treats members of (1) as counting for some enhancement of well-being and each higher kind of belief and intention as counting for more. The second account treats only some subset of the members of sets (1)-(4) as counting for any enhancement of well-being. He prefers the second option and suggests, but does not argue for, an account that treats only members of (4) as counting towards one’s perfection (Ibid, pg. 112-113).

There is yet a further aspect of Hurka’s theory. He thinks that even within the members of (4), there can be variations in the amount of perfection attained. He says that there is a *quality* to each member of (4) that must be taken into account (Ibid, pg. 114). Those beliefs and intentions with the greater quality, according to Hurka, are those that are general. The generality of a belief or intention, on Hurka’s account, is set by the breadth of the state of affairs it describes and, by its role within a hierarchical and explanatory system (Ibid, pg. 115).

I will turn now to a critical discussion of Hurka’s account. I am ambivalent as to whether Hurka’s account of theoretical perfection is plausible. However, I am certain that even if he has the correct account of perfection, human perfection could not serve in an account of well-being in the way he supposes. There are some problems that seem particular to Hurka’s favored version of human nature which do not necessarily apply to

all human nature accounts of well-being. There are more central problems, as well, which seem to be significant obstacles to Hurka's main arguments for the human nature account of well-being. I will explore these sorts of problems in turn.

Hurka's account seems too narrow in that it excludes too many beliefs from increasing one's well-being. The countless generations who believed, falsely, in some unifying scientific principle had a low level of theoretical perfection thereby, on Hurka's theory. It is not clear what principled reason there is for excluding *justified*, though false, beliefs in some contexts. A somewhat related problem is that people who generated what we now know to be false scientific theories, such as Newton, count as less theoretically rational than many who come after him, many of whom were much less good at science, even though Newton sure seems like a paradigm of the rational individual.²⁸ Theoretical rationality seems to involve more than just having beliefs of the right sort.

Now, Hurka could just expand his account to include merely justified beliefs as enhancing well-being. And he could even expand the account to include all true beliefs as enhancing well-being. Hurka does not give an explanation of why he prefers to restrict the set of beliefs and intentions to deserving successes, so it is hard to formulate a specific criticism on this issue. If Hurka were to expand the circle of which beliefs and intentions are relevant for well-being, it is not clear that there would not be new problems. Also, Hurka could change his position to include more than the mere holding of beliefs of the right sort in his account of theoretical perfection. It is because there seems to be some room to maneuver, that I do not think that these problems are necessarily damaging.

²⁸ If Newton is a poor example for some reason or another, then there are many others who could serve as examples.

Here is a criticism that hits Hurka's account more centrally. Hurka's account of well-being, given the role of human nature as described by Hurka, seems to fail to meet the necessary conditions of prudential well-being. Recall that interests that are relevant to one's prudential well-being are those that are non-moral, non-aesthetic, and self-involving. It is hard to see a principled way of ruling out counter-intuitive examples of beliefs that, on Hurka's account, end up counting as relevant to one's well-being. For example, knowing the number of blades of grass in Central Park, it turns out, increases one's theoretical perfection on Hurka's account. Now, the belief may have very little quality, to use Hurka's terminology, and thus might not increase one's amount of perfection by very much. But, it is an utter mystery how a true justified belief about the blades of grass in Central Park, in itself, could increase one's well-being at all. The belief fails to be sufficiently self-involving or self-regarding to count as relevant to one's well-being. It is not that having the beliefs under discussion decrease one's well-being, but these beliefs simply do not contribute to one's well-being.

Much of my criticism of Hurka has focused on his account of theoretical rationality, so let me turn to a problem with Hurka's account of practical rationality. As I say in the first chapter of this dissertation, any account of well-being must be able to explain acts and intentions of self-sacrifice. Hurka, given what he says about practical rationality, cannot explain self-sacrifice. The intention to spend one's entire life feeding the starving only increases one's well-being, when, in Hurka's language, it is a successful and justified attempt. That is, on Hurka's account, so long as one's plan to feed the starving works and one is justified in believing that the plan will work, one's perfection,

and so one's well-being, is increased. But this is false.²⁹ This would be an instance of self-sacrifice, not a way of enhancing one's own well-being.

It is hard to see how accounts of theoretical and practical rationality, given the assigned role of rationality in Hurka's theory of well-being, are sufficiently self-involving to serve as accounts of prudential well-being. With theoretical rationality, it is hard to see why merely having beliefs of the right sort intrinsically enhances someone's prudential well-being. Having true and justified beliefs will often lead to increases in prudential well-being—that is, instrumentally. After all, if one has good means/ends reasoning, one is more likely to get what one is after. However, on Hurka's view, merely to have such a belief itself brings about an increase in well-being, which seems implausible. With practical rationality, Hurka's inability to explain self-sacrifice is very problematic; he is committed to saying that all of one's successful and justified projects are in one's self-interest. This goes directly against the intuition that there are acts of self-sacrifice.

I think that the story of Frankenstein's monster is a good illustration of my criticism of Hurka's account of well-being. With this example, I hope to do two things: (a) show at least some reason for thinking that beliefs, as such, are not relevant to well-being and (b) show that there are other states of mind, loosely speaking, that are much more relevant to well-being. As I mentioned in the brief account of the novel above, after the monster's creation, he learns a great deal very quickly. To use Hurka's language, the monster increases the size of his set of deserving successful belief and intention attempts

²⁹ This problem comes from Hurka's choice to treat only "formal" considerations as relevant to the restriction of the set of beliefs and intentions that increase one's perfection. "Formal" criteria, according to Hurka, cannot make reference to the content of the belief and intention (Ibid, pg. 114).

very quickly. According to Hurka's account, his well-being should increase dramatically during this period. And yet intuitively, this is not the case. Throughout the time of learning, the monster is simply wretched. Here are some reflections by the monster on the matter:

I was not even of the same nature as man. I was more agile than they and could subsist upon coarser diet. . . . My stature far exceeded theirs. When I looked around, I saw and heard of none like me. Was I then a monster, a blot upon the earth, from which all men fled and whom all men disowned? I cannot describe to you the agony that these reflections inflicted upon me. I tried to dispel them, but sorrow only increased with knowledge. . . (Ibid, pg. 139)

There are many other instances in which the monster expresses hatred of his own existence. He likens himself to Satan in one passage, because he has been "cast down" by his creator, and then claims that he is worse off than Satan in another, because he has been left alone whereas Satan has some company. The kind of anguish he experiences evokes a great deal of sympathy. He states: "Of what a strange nature is knowledge! It clings to the mind, when it has once seized on it, like a lichen on the rock. I wished sometimes to shake off all thought and feeling, but I learned that there was but one means to overcome the sensation of pain, that was death" (Ibid, pg. 140). The monster is clearly suicidal.

During this period, the monster is physically healthy and develops his theoretical and practical rationality significantly. But, intuitively, his life is simply not going well for him. Now, to be fair to Hurka, the low level of the monster's well-being is not intrinsic to the new beliefs, but instead, is caused by those beliefs. The low level of well-being is caused by the beliefs under discussion because his holding and reflecting on these beliefs causes him great deal of unhappiness. So I think it would be incorrect to say that the monster's life is going poorly simply in virtue of having the "successful" and

“deserving” beliefs. But Hurka’s account is committed to claiming that the monster’s life is going quite well, and this just does not seem to be the case. This example is not conclusive, I understand, but it does seem compelling.

The above discussion raises the question as to what makes the monster’s life go poorly, at least intuitively. Either his unhappiness or his frustrated desires seem to be two good candidates that could figure into an explanation of why, intuitively, his life is going poorly. He is living his life in accordance with Hurka’s account of nature and yet has a poor life. Some other theory must explain the monster’s low level of well-being.

Hurka’s account of human nature is not the only explanatory objective account out there. Perhaps another account could do better. One of the reasons I work with Hurka’s account is that it is the best account of human nature I have found; both in terms of how well-developed it is and in terms of how good the arguments are for it.³⁰ So I do not think that any other account of human nature could serve better for the perfectionist model.

2.2.3 Application of Happiness to the Story

We turn now to a darker part of the story of Frankenstein. Though the above may evoke sympathy for the monster, perhaps what follows will not. The monster, after being rejected by everyone he comes across, and after having his hopes of having a spouse dashed, turns on his creator and makes the ruin of Frankenstein the goal of his life. Unbeknownst to Frankenstein, the monster is present when he tears apart the still lifeless body of his monster’s spouse-to-be. As the two part ways, the monster says to Frankenstein “I shall be with you on your wedding night” (Ibid, pg. 202). Frankenstein

³⁰ Note as well that I exclude evolutionary accounts of human nature from consideration at this stage because they could not plausibly serve in a theory of prudential well-being.

takes this to be a threat against him; that is to say, Frankenstein thought the monster was threatening to kill him on his wedding night. However, the monster's intention was to kill Frankenstein's new bride, Elizabeth.³¹

Not wanting his wife to see a fatal battle between himself and his monster, Frankenstein sends his wife to their bedroom. Frankenstein wants her to be safe, but she is not. He is very nervous and agitated, but is not unhappy. He does not fear for the life of his new bride. The monster sneaks into their bedroom and kills her. Frankenstein finds out very soon thereafter that she has been killed.

Frankenstein, in recollecting this sad tale, remarks: "Great God! If one instant I had thought what might be the hellish intention of my fiendish adversary, I would rather have banished myself forever from my native country, and wandered a friendless outcast over the earth, than have consented to this miserable marriage. But, as if possessed of magic powers, the monster had blinded me to his real intentions and when I thought that I had prepared only my own death, I hastened that of a far dearer victim" (Ibid, pg. 231).

Frankenstein's life goes worse when his new bride's life is threatened and even worse when she is killed. But the mental state account does not account for this.³² It has it that Frankenstein's life goes worse when he "finds out" about Elizabeth's death. The gap between her death and Frankenstein's finding out about her death is extremely short (the time it takes for Frankenstein to run to their room). But there are larger gaps of time in the other murders. Days, perhaps weeks, in the case of one of Frankenstein's relatives

³¹ No doubt, the monster sees this as a more fitting punishment for Frankenstein's decision to abort the creation of another being to be the monster's wife.

³² A newer kind of mental state account, such as that of Sumner (1996), requires that happiness and unhappiness be *informed* to count toward or against one's well-being. Sumner's account is what I would call a "mixed theory," and discussion of it must be put off till later.

who is killed by the monster. Of course it is conceivable that Frankenstein would never find out about some of the murders. Yet the mental state accounts say that Frankenstein's life would go well so long as he never found out about them. The objective account and the desire account have something going for them here that the mental state account lacks: namely, that there are cases where peoples' lives go poorly for them despite the fact that they are not aware of this. I do not take this point to be conclusive against the mental state conception, but it does show a weakness.

For a more telling objection, I must modify the story slightly. In Shelley's story, Frankenstein's creation gets angry when he is rejected and gets depressed when he feels utterly alone and isolated. He lashes out at Frankenstein and all of society in his rage. Frankenstein wants to pacify his monster and so, let us say, to modify Shelley's story, he creates an experience machine. The experience machine generates experiences for the monster such that he will have whatever mental state a mental state theory takes to constitute well-being. This idea, of course, is drawn from Robert Nozick's famous thought experiment found in his *Anarchy, State and Utopia*. Nozick frames the issue as one about whether people would choose to go into the machine. This element of choice needlessly complicates the matter. I will make the story such that Frankenstein just places the monster in the machine without the monster's knowledge. We must think about whether the monster's life is going better when he is in the machine.

Once in the machine, Frankenstein's monster has whatever mental state is relevant to well-being according to any given mental state conception, whether it is pleasure, happiness or some third alternative. The monster can be made to believe that all of his wishes are granted and he could turn out quite content and happy. Has the monster's

well-being been increased thereby? It is implausible to think that merely putting the monster in the machine increases his well-being.

Nozick offers three reasons for thinking that the mental states which result from living in an experience machine could not be all there is to well-being; life in the experience machine leaves something important out of the life. The first reason is that people want to *do* things rather than just experience them (Nozick, pg. 43). The experience machine does not distinguish between doing something and experiencing it. The second is that people want to *be* a certain way rather than just to experience things (Ibid, pg. 43). It is a bit difficult to tell what Nozick has in mind with this second idea, but the way he expresses the concern is that, once inside the machine, we will not be any particular sort of person (Ibid, pg. 43). The third reason is that living in the machine limits one's "reality" to a man-made "reality" and people want to be connected to the deeper reality (Ibid, pg. 43-44). Nozick suggests that once in the machine, there is no "actual contact" with any deeper reality (Ibid, pg. 43). Perhaps Nozick has misstated the problem just a bit; there could still be *some* connection to a deeper reality when in the machine. But even if there is some connection to a deeper reality, it is less substantial than we would prefer. So I think Nozick's third point needs to be modified a bit, but still maintains much of its original force.

Nozick's three points about the inadequacies of life in the machine are generally quite convincing. A person's life in the machine, even Frankenstein's monster's life, would be so separated from his "mental life" that his life cannot plausibly be said to be going better. There must be, as it were, some change in the world external to his

psychology in order for his life to go better at least some times.³³ There might be some cases where we might think that something that happens in the experience machine does make one's life go better or worse, but that certainly will not be the case across a full range of examples. With the mental state theories, no such external changes are necessary; thereby making the theory counter-intuitive.

Now the experience machine poses less of a challenge to the theories that base well-being in human nature or desire-satisfaction than to mental state accounts. The monster's beliefs will be false and his intentions will rarely succeed and so, according to Hurka's account of which beliefs matter for perfection, the monster's well-being would be low. As I mentioned earlier in this chapter, Hurka mentions other accounts of which beliefs matter for perfection (such as justified beliefs) and so there might be some flexibility in the matter. My thinking on this is that so long as there is flexibility in the issue, human nature accounts will be able to account for experience machine cases with at least a minimum of adequacy.

Desire-satisfaction accounts do not imply counter-intuitive results in experience machine examples. Now, some desires can be satisfied within the experience machine, such as desires to have certain experiences. So according to at least many desire-satisfaction theories, one's well-being could increase when the experience machine produces the desired pleasure. However, I do not think that these sorts of examples are counterintuitive for the desire-satisfaction theories.

Here is just a simple example meant to illustrate a point about desire-satisfaction theories and experience machines. Imagine that an agent wants to experience the taste of

³³ In a later chapter, I will examine a theory that combines mental states with facts outside of the mental life of the agent. Such a theory is a mixed theory rather than a pure mental state theory.

mint chocolate chip ice cream and, lacking any ice cream of that sort, opts for the experience machine. The agent's well-being could thereby increase according to desire-satisfaction accounts. Those who think such examples problematic for desire-satisfaction accounts of well-being need to pay careful attention to what the agent's desires are and which desires are satisfied. Compare the content of three closely related, but different, desires: (1) to experience the taste of mint chocolate chip ice cream, (2) to taste mint chocolate chip ice cream, or (3) to eat mint chocolate chip ice cream. Of these three desires, only the first could be satisfied within the experience machine, as the machine is usually understood. Strictly speaking, the second and the third could not be satisfied when in the machine. The second and third desires cannot be satisfied because the person does not actually taste any ice cream, nor does he actually eat ice cream, when in the experience machine as it is normally understood. Although the example is simple, it illustrates something about the content of our desires. Though people often speak loosely about the content of their desires, careful attention to the content of desires is necessary when dealing with such cases. Very close attention to the content of the desires in these cases makes the experience machine cases unproblematic for the desire-satisfaction account of well-being.

The three problems that I mention Nozick thought the experience machine posed for certain theories do *not* show that all accounts of well-being that treat well-being as increasing in the machine are problematic. The three problems only show that any theory which treats well-being as consisting entirely in states produced by experience machine is inadequate. So the experience machine shows that mental state theories are inadequate. But desire-satisfaction and human nature theories have no special problems dealing with

experience machine cases. We can allow for some increases in well-being, according to a theory of well-being, within the experience machine without implausible consequences for that theory. We just have to be able to say what else makes a life go well besides mental states.³⁴

So, there are serious problems with pure mental state accounts. Later in this dissertation, in Chapter 6, I will examine more sophisticated theories of well-being that get around the experience machine cases. But for now, note that desire-satisfaction accounts seem to avoid the objections to which mental state and explanatory objective theories fall prey. In the next chapter I will develop what I take to be the most plausible desire-satisfaction account. In 2.2.4, however, I will examine a different kind of account of well-being; an account which does not pick out just one thing that makes a life go well, but instead, picks out many different things.

2.2.4 Pluralistic Conceptions of Well-Being

An account is pluralistic if it treats two or more factors as each being partly constitutive of well-being. With pluralisms, each factor is a well-being “maker;” i.e. each factor is at least partly *constitutive* of well-being—the presence of each enhances well-being. Pluralisms are not to be confused with *mixed* theories. A mixed conception takes several different factors and adds them together as **necessary** conditions. Sumner’s view, discussed briefly above as proposing a combination of mental states and states outside one’s mental life to constitute well-being, is a mixed theory. Pluralisms differ from

³⁴ Nozick himself states the three problems for mental state theories stemming from the experience machine cases in a way that is very friendly to desire-satisfaction theories. I paraphrase him above, but I stay true to his language of putting the objections in terms of what we want for ourselves. For example, Nozick states, roughly, that we want to do certain things and be certain sorts of people (Ibid, pg.43). Desire-satisfaction accounts have no trouble including these desires in the account of well-being. Indeed, they actually account for the missing elements quite well.

mixed theories in that each factor is **sufficient** for at least some amount or degree of well-being. Assessing mixed theories can be quite complicated because the mixed theories start with one of the above discussed conceptions but then change it in a quite fundamental manner, thereby losing either one or both of the characteristics that make it plausible in the first place. I must put off discussing mixed theories until a later chapter in this dissertation because they are so sophisticated. When one draws up a list and does not unify the items on the list, one way of understanding the list is in the pluralistic way.

John Finnis, in his *Natural Law and Natural Rights*, offers a list of things as an account of “‘good,’ ‘basic good,’ ‘value,’ ‘well-being’” (Finnis, pg. 86). As I said before, Finnis lists the following factors as relevant to well-being: life, knowledge, play, aesthetic experience, sociability, practical reasonableness and religion. Griffin has a similar list.

Finnis claims that the fact that these element belong on his list is “indemonstrable” and “self-evident” (Finnis pg. 59, 65, and 85). Sometimes views like Finnis’ are dismissed quite quickly. For example, Kitcher in his essay “Essence and Perfection” states:

Distinguish two varieties of objectivism. *Bare objectivism* simply offers a list of the things that make human lives go well. When asked what qualifies the items for inclusion, bare objectivists have no explanatory theory to offer; it’s simply a brute fact that these things are good for us . . . Clashes between bare objectivists seem doomed to immediate stalemate. (Kitcher, pg. 59-60).³⁵

I think that a lot of the criticism of so-called “objective list” theories is due to the claim their advocates often make that such lists are *self-evident*. Other authors disagree with them and do not like to be told that their “faculty of perception” of well-being is

³⁵ See also Sumner’s *Welfare, Happiness, and Ethics* (pg. 24, 46).

malfunctioning. However, there is no need to accept Finnis' claim that his list is self-evident.

The items on the list, then, should be thought of as having an epistemic status relevantly like the status of the set of mental states, or satisfactions of the desire set, or the aspects of human nature, in other kinds of theory.

So I propose to read Finnis as a pluralist and ignore his claim to self-evidence. Griffin's latest book, *Value Judgement*, contains a somewhat similar list that can also be interpreted as pluralistic. These theories are to be evaluated just as the other theories are, by engaging in critical reflection—looking at the implications of the theory and then reexamining the theoretical underpinnings of the conception.

These pluralistic accounts look very attractive at first because they seem to solve at least some of the problems that the other theories face. In fact, though I do not know of anyone who does this, one could generate a pluralistic theory that combines the factors referred to as enhancing well-being in each of the three types of monistic conception—some mental states, desire satisfaction, and whatever accords with human nature.³⁶ Generally, pluralists offer six or seven items on a list with the claim that no item on the list is reducible to any other and that none is more important than any other.

Let us just look at one example where it seems that a pluralist can fix the problems that beset an explanatory objectivist such as Hurka. Think back to the example of Frankenstein's original monster in the unaltered story. Frankenstein's monster is

³⁶ Amartya Sen in his "Plural Utility" discusses the possibility of a consequentialism that treats utility as composed of both mental states and desire satisfaction in a pluralistic fashion. Scanlon comes close to accepting a theory that treats mental states, desire-satisfaction and other features as each constituting at least some well-being (Scanlon, pg 124-125). However Scanlon claims that his account of well-being is still incomplete in that it does not give a final way of deciding between conflicts of the various factors that make a life go well; it is not a *theory* of well-being in the full sense (Ibid, pg. 125).

developing his natural talents but is nevertheless extremely miserable. A pluralist could include some subjective elements to account for our intuition that the monster's life is going poorly. The fifth element on Finnis' list, sociability (friendship), is missing from the monster's life. In this way, it seems that a pluralist could account for some of the counterintuitive results of an explanatory objective theory.

There is one complication here: it appears that the pluralist who endorses Finnis' list must claim that the monster's life is going *quite well* because he has nearly every element on the list (and in high amounts), but that his life would go *better* if he had social relations. This is counterintuitive, because it looks like the monster's life is going very poorly. What this might mean is that Finnis' list is suspect. However, it is open to the pluralist to change the list to make it more plausible. So this is not an objection to pluralists *per se*. One could just draw up a list containing mental states, desire-satisfaction, and perhaps friendship.

So pluralistic theories, it seems, have some plausibility in virtue of their inclusiveness. If, intuitively, there is something missing in someone's life that thereby makes the person less well-off, one could add whatever is missing to the list. However, this permissive inclusivity, which at first seems like a great virtue of this kind of theory, turns out to be a significant liability. Let me explain.

Any *monistic* conception of well-being picks out just one factor and says that well-being varies with that factor, no matter how complexly that factor is characterized. As I have already mentioned, mental state, desire-satisfaction and explanatory objective theories are all monistic. *Pluralistic* theories, on the other hand, pick out two or more things. For the pluralist, some actions will enhance one factor but diminish another

factor. For example, going to church on a regular basis may increase the “religiosity factor,” but decrease the amount of play in one’s life. Both play and religion are on Finnis’ list. What the pluralist needs is an account of the weighing and measuring of the various different factors that make a life go well in order for the theory to give definite assessments of well-being—a principle of adjudication that makes commensurable the factors that constitute at least some well-being.³⁷ The principle of adjudication would say when one’s life is going better than, worse than, or equally well as the life of another person (or perhaps another time-slice of the same person).

There are two sorts of problems for the pluralist resulting from the lack of a principle of adjudication; metaphysical problems and epistemological problems. The metaphysical problem for pluralistic theories is that without such a principle, there is no fact of the matter about how much well-being someone has when she has two or more of the factors constitutive of well-being. The epistemological problem for pluralistic theories that arises from the absence of a principle of adjudication is that we have no way of knowing how much well-being one has when she possesses two or more of the features that could contribute to one’s well-being.

Now, I think that the metaphysical problem is much worse than the epistemological problem, but both seem to be significant problems. In everyday judgments, people assess levels of well-being and make choices in cases of tradeoffs. What the pluralist needs to do is to vindicate the idea that there are really are amounts of well-being and so solve the metaphysical problem. But we also would very much like a philosophical account of what epistemological tools one could use to make such judgments. However, it is

³⁷ One interesting issue is whether the principle of adjudication will then turn what looked like a pluralistic conception into a monistic conception. This is an interesting issue, but too far off my course.

because the epistemological problem is not unique to pluralistic theories that I think it is less significant problem.³⁸

Thus far, no principle of adjudication has been given and none seems forthcoming. Suppose that the pluralist admits that a principle of adjudication cannot be given and claims instead that intuition can solve the problems. Intuition, on this account, could then *determine* the amount of well-being someone has (to solve the metaphysical problem) or intuition could at least serve as a guide in *discovering* or detecting the level of well-being one has (to solve the epistemological problem). The appeal to intuition will certainly not work to solve the metaphysical problem and it probably will not work to solve the epistemological problem. People differ so widely in their intuitions on the same cases. There is not a clear reason for thinking anyone's intuitions are superior and even if one were to grant that there could be authorities in such an issue, there is not a clear reason for picking out the authority.

The pluralist might grant that there is no principle of adjudication, and that relying on individual intuition will not work. He could simply admit that there is a radical incommensurability involved. This incommensuration is especially troubling when viewed as metaphysical incommensuration. There simply would be no fact as to how much well-being someone has when she possesses two or more features on the list of things that make a life go well. This option amounts to the claim that people in some cases do not have any definite total amount of well-being at any given time. The pluralist

³⁸ All of the monistic theories discussed in this chapter have at least some epistemological problems. For example, a proponent of Hurka's perfectionist theory would no doubt have problems identifying a person's entire belief set: the epistemological problems for Hurka's perfectionism are much worse if trying to determine which beliefs are true or false. As I say, I believe other monistic conceptions have epistemological problems too.

of this sort admits that there are many cases where there is no answer to questions about whether one's life is going better, worse, or remains the same. But this is problematic, for then our conception of well-being fails to live up to our ordinary judgments about well-being.³⁹ This kind of incommensurability is of a different order that is, say, *interpersonal* incommensurability. This kind of incommensurability is what could be called *intrapersonal* incommensurability. At least in many cases, it would make it impossible for there to be any fact of the matter as to whether someone's life is improving or getting worse overall. All that the pluralist could say about some cases is that "insofar as *A* has feature *x*, *A*'s life is better; but insofar as *A* has *y*, *A*'s life is worse." Intuitively, we make ordinary judgment about overall well-being and think that there is a fact of the matter regarding overall well-being. Pluralism that accepts incommensurability must reject such ordinary, everyday judgments.

Pluralisms are inclusive in that they bring in to the theory many components that can make a life go well. The cost, however, is that pluralisms result, in a way, in a sort of fragmentation of well-being. If all one can say is that one's life goes well in virtue of *x*, but poorly in virtue of *y*, then the overall assessment of a life that should be intuitively fairly clear, at least in many cases, is missing.

We have looked at two methods of making the items on the list commensurable; using a principle of commensuration and using intuition to commensurate. No principle of commensuration is forthcoming and there is no account of whose intuitions are

³⁹ We have been looking at the difficult cases lately. But I ask the reader to go back to the "easy" cases; those discussed with the Count of Monte Cristo or commonsense everyday judgments. It is difficult to test one's intuitions in cases where the various conceptions of well-being differ. The examples from the first chapter are better (in that one's intuitions in such cases should be clearer) because all of the plausible conceptions of well-being "agree" about what to say regarding well-being in those cases.

authoritative. If the pluralist accepts incommensurability of the factors on the list, then the pluralist cannot vindicate ordinary judgments of well-being in simple cases, either metaphysically or epistemologically.

Perhaps there may be a way to save pluralism, however. There has been an intriguing proposal by Ruth Chang in her “Introduction” to *Incommensurability, Incomparability, and Practical Reason*. Chang claims that there is a fourth positive value relation in addition to “worse than,” “equal to,” and “better than.” This fourth relation can be described as “on par with” (Chang, pg. 27-28). The “on par with” relation is “positive” in the sense that it is a kind of comparability as opposed to incomparability. If two things, ϕ and ψ are on a par, and a third thing, $\phi+$, is better than ϕ , there is still room for $\phi+$ and ψ to be on a par as well (Ibid, pg. 24). The “on par with” may seem paradoxical in some ways. Chang does not think that the “on par with” relation is a kind of vagueness and so cannot mean “roughly equal to” or anything similar. Furthermore, judgments involving the “on par with” relation are not merely those involving ignorance about the “true” value of the things in question. Chang’s “on par with” relation is supposed to be a relation in the metaphysics of value and so looks like it might go some way to solving the metaphysical problems that pluralisms face.⁴⁰ Whether a fourth positive value relation of the sort Chang has in mind exists is controversial. Chang’s proposal can be argued against on several fronts. Rather than go into the arguments about whether there indeed is a fourth positive value relation however, I would like to

⁴⁰ If a pluralist adopted Chang’s “on par with” relation, perhaps some of the force of the epistemological problems could be lessened as well. For, if there is no concrete fact of the matter that one has more or less well-being than another in some situation, we should not demand that any theory have an account of how one is to, or at least could, decide such cases.

explore the prospects of using the “on par with” relation to solve the problems of commensurability for the pluralist about well-being.

Chang’s “on par with” relation seems, at first glance, like it might be able to solve the incommensurability problem pluralisms face.⁴¹ What a pluralist could say is that the amounts of well-being enhanced by having some of either of two items on a list are on par with each other. For example, the pluralist could say that, other things being equal, the level of well-being of someone with a minimum amount of religiosity is on par with the level of well-being of someone with a minimum amount of friendship.⁴² The pluralist could claim that many lives are on par with each other.

Adopting Chang’s on par relation does not really help, however, in that it does not solve the *fundamental* problem for the pluralist about well-being. For, even if we treat lives that have minimal amounts of each of the items on the list as being on par with each other (other things being equal), the deep metaphysical and epistemological problems remain. There is still nothing about a pluralistic theory that could, *even in principle*, solve the remaining metaphysical and epistemological problems. Let me explain. Presumably for an account of well-being to be adequate, there must be a fair range of cases where one life is worse than, better than, or equal to another life in regards to well-being. Even with Chang’s proposal in place, the pluralist has nothing in his theory that

⁴¹ I should note that Chang does not offer the “on par with” relation as a solution to the problems that the pluralisms about well-being face. Instead, I have taken what she says about two things being on a par and will examine whether it can help the pluralist out of his problem with incommensurability.

⁴² Interestingly, perhaps the applicability of the “on par with” relation could be extended. Perhaps a pluralist could say that an agent’s well-being at a time is on par with his past level of well-being, when the person has at least some of the two different factors that make a life go well and one changes over time. For example, at time t_1 , A has a great deal of friendship and a little bit of religiosity; later at t_2 , A has the same amount of friendship but a bit more religiosity. A’s level of well-being at t_1 and t_2 could be on par with each other. I will not explore this possible extension.

could settle the matter except for in a few rare instances. The pluralist who accepts Chang's "on par with" relation would still need a sort of principle of adjudication or a sort of intuition to determine when two lives are on par. The pluralists, even if he accepts Chang's "on par with" relation, must still provide an account of weighing and measuring for many of the ordinary lives. This is a tall order, perhaps not as daunting a task as the fully precise principle of adjudication or intuition, but nevertheless an extraordinarily daunting task.

Let me illustrate how the central problem still remains for a pluralist who adopts Chang's "on par with" relation. Suppose someone has one and only one item on the list, say, friendship. According to a simple version of pluralism with the "on par with" relation, his life will be on par with many other possible lives, such as the possible life in which he has a great amount of pleasure. Both possible lives will be on par, according to the simplistic theory under discussion, because the person in the examples has a bit of each of the items on the list. Too many lives will be on par according to a pluralist who adopts the on par relation. Perhaps, by adopting the "on par with" relation, a pluralist could explain how one life with a little bit of an item on a list is worse than a life with more of the same item, but the problem remains about how to distinguish between lives that have differing, perhaps many, items on the list. A principle of adjudication is still necessary. Perhaps with the "on par with" relation, only a rough principle of adjudication is necessary. But even a rough principle is a very tall order.

At first glance, pluralisms look attractive. It seems like they may be able to solve the problems the various monisms face. Their inclusiveness is very appealing initially. However, upon further examination, they come with a set of insurmountable problems of

their own. In the absence of a method that makes the items on the list commensurable, whether by principle of adjudication (which does not seem forthcoming) or intuition (which does not seem up to the task), they have striking difficulties dealing, in a philosophical way, with seemingly easy judgments of well-being. The pluralist who adopts Chang's "on par with" relation does not solve the central problem.

2.2.5 Conclusion

Desire-satisfaction accounts of well-being seem to come out ahead of its competitors at this point of the critical discussion of conceptions of well-being. Explanatory objective theories have a hard time excluding the evolutionary account of human nature, which could not plausibly serve in an account of prudential well-being. Non-evolutionary accounts of human nature, such as Hurka's, also face serious problems. His account of physical nature has at least minor problems, but the bigger problems are with his account of rationality. Developing and exercising one's theoretical and practical rationality fails to be self-involving enough to meet the conditions of the concept of well-being.

Mental state theories have problems too. Their initially attractive feature, the dependence on the psychology of the agent, turns out to be an enormous liability when we think of actual and hypothetical cases where one's psychology can be easily manipulated. Thought experiments involving the experience machine illustrate these problems quite clearly and convincingly.

Although I think that pluralisms have not been sufficiently explored and appreciated, they come with huge problems of their own. Given that no principle of commensuration has been forthcoming, there is not any obvious way to vindicate many ordinary judgments of well-being. Adopting Chang's idea that two possible lives can be

on par when two or more items on the list are had by the agent—or the idea that two possible lives can be on par when an agent has some amount of two different items on the list—is initially intriguing, but leaves us no better off in assessing ordinary lives; for most lives have multiple items on the list to varying degrees and, at least in many cases, we have straightforward intuitions on such matters.

To be fair, I have only shown how the desire-satisfaction account is not susceptible to the problems of the other conceptions. I have not addressed problems that are specific to desire-satisfaction accounts. I will do that in chapters 3-6, where I develop and defend what I take to be the best version of the desire-satisfaction account.

CHAPTER 3 DEVELOPING A DESIRE-SATISFACTION CONCEPTION OF WELL-BEING

3.1 Actual Desire Accounts Critically Discussed

3.1.1 Some Introductory Remarks

There is a lot to be said in favor of desire-satisfaction accounts of well-being. We have seen some of their plausible features in play in the previous chapter. Before we move on to discuss the more sophisticated mixed views of well-being, it is necessary to develop the most plausible version of the desire-satisfaction conception of well-being. Let us use the term “relevant desire set” to refer to the set of desires the satisfaction of which increase well-being. Some may say, somewhat implausibly, that the relevant desire set for an agent includes all actual desires of that agent. Most desire-satisfaction theorists have thought that the relevant desire set is something different from the set of total actual desires. The project of this chapter is to identify the relevant desire set. The project of the next chapter (Chapter 4) is to defend that account against recent criticisms. The project of Chapter 5 is to make sure that the desire-satisfaction account developed and defended in the third and fourth chapters meets the conditions of prudential well-being as set out in the first chapter.

So this chapter is dedicated to identifying the most plausible set of desires the satisfaction of which increase well-being. It is because the desire-satisfaction account will be modified slightly in the third and fourth chapters that the reader should think of the theory developed in this chapter as a “rough draft.” The final version of the central theory should be made clear by the end of Chapter 5, but we must start somewhere. In

this chapter I will first examine whether the relevant desire set ought to be restricted to intrinsic desires or whether the satisfaction of instrumental desires ought to count in favor of one's well-being as well. I conclude that restricting the desire set to intrinsic desires makes for a more plausible account of well-being. Then I will go on to discuss whether we should think of the relevant desire set as being composed of actual intrinsic desires or intrinsic desires one would have in counterfactual situations. I conclude that intrinsic desires one would have under counterfactual conditions (which I will specify) are best.

3.1.2 Intrinsic and Extrinsic Desires

As mentioned above, I will argue that only intrinsic desires are relevant to well-being. Before I get into the arguments, it makes sense to provide an account of intrinsic and extrinsic desires. It is commonly said that intrinsic desires are desires for things in themselves. This common way of understanding intrinsic desires will do for now, though it will be made clearer after the discussion of extrinsic desires below. No desire can be intrinsic and extrinsic; things can be intrinsically desired and extrinsically desired, but no desire can be both intrinsic and extrinsic.

I will discuss three types of extrinsic desires: (1) instrumental desires, (2) whole-constitutive desires, and (3) part-constitutive desires. A desire is instrumental if and only if one has it because one believes that the satisfaction of it will be a means to some end. For example, Frankenstein may have a desire for money. This would be an instrumental desire if he has it because he believes that the satisfaction of it will help in satisfying his desire to create his monster.

Secondly, a desire is whole-constitutive if and only if one has it because one believes that the satisfaction of it constitutes the satisfaction of another desire entirely. For example, Frankenstein may have a desire to bring a corpse to life because he believes

that in doing so he will do something God-like. The satisfaction of Frankenstein's desire, he believes, simply constitutes doing something God-like. Let me return briefly to intrinsic desires. It may seem, at least at first glance, that whole-constitution desires are intrinsic. This is a mistake. If Frankenstein really desires to bring a corpse alive *because* he believes that in doing so he will do something God-like, then Frankenstein's desire to bring a corpse to life makes reference to some other desire. If Frankenstein desires something because of some other desire, then the desire is extrinsic.

Thirdly, a desire is part-constitutive if and only if the agent has it because he believes that satisfaction of it partly constitutes the state of affairs that is the object of another desire. For example, Frankenstein may have a desire to have a brain (not the brain in his head, mind you, but another brain). He has the desire for a brain because he believes that having it will achieve part of having a whole corpse. Strictly speaking, it is not that the brain is a means to having a corpse: rather, it is that having a brain is part of having a corpse. Frankenstein's desire for a brain would be part-constitutive in this case.

I wrote above that an intrinsic desire, roughly, is a desire for something in itself. An agent has an intrinsic desire for something, on my account, if the agent desires that thing without some other desire playing a role as a reason for the desire. Now, an agent may have an intrinsic desire that her desires are satisfied; so her intrinsic desires might make reference to desires. But she cannot have an intrinsic desire *because* of some further desire.

There is one last thing to note about intrinsic and extrinsic desires. It seems entirely possible that a desire could begin as an extrinsic desire and then change to an intrinsic desire. Frankenstein's desire for money, which begins as an instrumental desire

could change to a desire he has intrinsically. The “because” in the above formulations is not supposed to capture something about the generation of a desire. Rather, the “because” should be thought of as invoking a reason or justification from the agent’s point of view. So, even if someone desires something in itself and she has the desire “because” (in an historical way) she had an extrinsic desire of a particular sort, her desire should still be thought of as intrinsic.

Should the satisfaction of extrinsic desires count towards one’s well-being? As I have put it, should extrinsic desires be in the relevant set? I do not think that they should be in the relevant set and in 3.1.3 and 3.1.4, I argue for that conclusion.

3.1.3 Problems for Total Desire Satisfaction Accounts

Total desire accounts take the satisfaction of all one’s desires as increasing one’s well-being, whether extrinsic or intrinsic. As I suggest above, I believe that actual desires, at least in some cases should not be in the relevant desire set, but let me set aside the issue for now. On total desire accounts, both extrinsic and intrinsic desires matter. They do not have to matter equally. Desire sets can have a hierarchy. An outcome can be desired because it is instrumental to another outcome, which itself might be instrumental to some further desired state of affairs. Hierarchy matters on these accounts, insofar as there are differing “weights” that can be placed on the desires.¹ Some will be more important than others. Even taking hierarchy into account, I think that total desire-satisfaction accounts have significant problems as a theory of well-being.

Imagine Frankenstein is working in his laboratory (this does not happen in the book). He is tired and exhausted, but thirsty. Water will quench his thirst, he thinks. He

¹ There could be other forms of hierarchy as well—even between different intrinsic desires. At this point, I only want to focus on one sort of hierarchy as a resource for the total desire view.

reaches out to what he believes is a glass of water in front of him. He pulls it to his mouth and takes a drink. On this imagined example, he has an intrinsic desire to quench his thirst and an extrinsic desire to drink from the glass in front of him. Frankenstein's desire to drink from the glass in front of him is satisfied. Has Frankenstein's well-being increased as a result of his desire-satisfaction? Well, that depends, I take it, on whether the glass in front of him contained water or not. Let us suppose that it did—that seems a case where it is plausible and intuitive to say that his well-being has been thereby increased. However, let us suppose that it did not contain water. Not to expose the reader to an example that is too disgusting, let us say the glass contained alcohol. So Frankenstein really drank alcohol when he satisfied his desire to drink from the glass in front of him. It seems clear here that this cannot be a case of improved well-being, but read on.

We cannot put the total desire theory in the coffin just yet. Frankenstein's desire to drink from the glass in front of him is just one of his desires among many. He has other desires as well in the case where the glass contains alcohol. Maybe the satisfaction of this one desire has caused the others to be thwarted (such as the desire to be clear-headed). The person who favors a total desire-satisfaction theory could argue that, in the story so described, Frankenstein's well-being suffers a net decrease as a result of the action—so the example does not have counter-intuitive results. Call this response the “swamped desire” response. The desire to drink from the glass can be swamped by other desires both in number and in importance. But I think that the swamped desire response will not work, for it grants the person who favors total desire-satisfaction far too much. The result (after the net is calculated) comes out all right, but the way of getting to that

result is still unacceptable. The following thesis still holds for a total desire satisfaction theory: Insofar as Frankenstein drank from the glass in front of him, his well-being was improved. Given that it contains alcohol and not water, this is counter-intuitive. Additionally, the “swamped desire” response only works if one has enough desires (or enough strong desires) thwarted. There will be cases where there are no such countervailing desires and in those cases, an agent’s desire satisfaction will increase his well-being.

Frankenstein’s desire to quench his thirst seems to be relevant to his well-being. Furthermore, let us suppose that his desire to drink from the glass in front of him is derived from beliefs that are highly justified: Frankenstein has not been a careless reasoner. The problem then would lie in counting the desire to drink from the glass in front of him as relevant to his well-being. There are two ways to block this undesired consequence: (1) remain with actual desires, but restrict the relevant set to only intrinsic desires, and exclude extrinsic desires: this is to abandon total desire views or (2) to move away from an actual desires account to a counterfactual account of the desires, the satisfaction of which, improve one’s well-being: this is compatible with remaining true to total desire views. I will discuss these two strategies, or ways of responding to the objections, in turn.²

3.1.4 Intrinsic Desires

Restricting the relevant class of desires to intrinsic desires would solve the above-mentioned problem. On the story given, Frankenstein intrinsically desires to quench his

² Sometimes writers on this topic will look at problems structurally similar to the “thirsty Frankenstein” example and immediately conclude that the way to solve the problem is to go with (2), but this is to miss an important possible solution to this problem. For an example of someone who immediately goes with (2), see Griffin, 1986. I think it very probable that Griffin understands the deeper issues, but his example is misleading.

thirst. His desire to drink from the glass in front of him is merely instrumental to the end of quenching his own thirst. He has no intrinsic desire to drink from the glass in front of him. Even though he drinks from the glass in front of him, his well-being is not increased thereby. In fact, with plausible background values put into place, presumably his well-being is decreased by drinking alcohol because he has intrinsic desires to stay healthy, clear-minded, and so on.

It might seem one could solve the problem—alcohol—by restricting the desire set to exclude desires based on false beliefs—i.e. by going to hypothetical fully informed desires. Let us take a look at the hypothetical Frankenstein example again, but modify it slightly. Let us go with a case where there are true beliefs that link up the intrinsic with the extrinsic desire (the idea being here that we should look at instrumental desires when they are at their best and see whether satisfaction of them increases one's well-being). So in this newly described, but similar case, Frankenstein is thirsty, desires to drink from the glass in front of him, but this time, there is water in the glass in front of him.

Frankenstein drinks and both desires are satisfied.

I am far from sure that even on this newly described case, we should think that there is something to be said for the idea that the satisfaction of one's instrumental desires increases one's well-being. Frankenstein, so described, has two desires both of which are satisfied: the one to quench his thirst and the other one to drink from the glass in front of him. Does this mean that the successful completion of his action counts twice in favor of his well-being? That does not seem quite right. Suppose now that he has to move a sheet of paper that is on top of the glass and he desires to move the sheet of paper to drink from the glass. Does the satisfaction of all three desires count toward his well-

being? It seems the improvement in his well-being is the same in the case where he does move the paper and in the case where there is no paper to move.³ The driving intuition in all of these cases is that what matters is the quenching of the thirst, not grasping the glass, or moving the glass closer to his face or whatever. So, even if we look at instrumental desires in their best light (when they are correctly related to our intrinsic desires), satisfaction of them does not seem to improve our well-being.

We are looking for an account of the set of desires the satisfaction of which lead to an increase in well-being (the so-called “relevant set”), and total desire views do not seem plausible. This is so when we go with actual desires and even extrinsic desires that are correctly connected to the intrinsic desires. Now that the relevant set has been restricted to intrinsic desires, let us look to see whether one’s actual intrinsic desires can be in error and whether there is a way of correcting for this error in the relevant desire set by moving to a hypothetical desire set.

3.1.5 A Critical Discussion of Actual Intrinsic Desire Views of Well-Being

My immediate project is to critically assess actual intrinsic desire views of well-being. I will argue that actual intrinsic desire views are inadequate in two ways: they do not explain cases that involve ignorance in an intuitively attractive way and they do not explain our intuitions about cases where one would cease to have an intrinsic desire upon confrontation with facts. If actual intrinsic desires constituted the best relevant desire set, the resulting desire-theory of well-being would be relatively simple. However, and

³ There is an interesting slightly modified case where the glass of water is trapped in a box and Frankenstein would have to spend two hours getting it out of the box (there is no other way to quench his thirst). One might think that getting it out of the box constituted an achievement of sorts and ought to count in favor of his well-being. But I think that this is mistaken. I think that what is going on in the background is our presupposition that Frankenstein, like the rest of us, desires to achieve his goals with as little effort as is required or that he desires and enjoys exercising of ingenuity, etc.

unfortunately, actual intrinsic desires prove inadequate as constituting the relevant desire set. A desire-satisfaction theory can deal with the two lines of criticism that the actual intrinsic desire theory faces if it treats the relevant desire set as made up of the desires one would have under certain counterfactual conditions. I will leave the specification of the counterfactual conditions for later. Here, I will show that such a counterfactual specification is desirable by showing that actual intrinsic desire views are inadequate.

I think that actual intrinsic desire views fail to account for an agent's well-being when the agent is ignorant, at least in some situations. Let me explore a few hypothetical cases. Imagine a case in which an agent is faced with a choice between eating some pineapple and eating some watermelon. In this example, let us suppose that the agent has never heard of pineapple, much less tasted it. She has tasted watermelon and likes it, though not strongly. In this case, she desires to eat the watermelon and she does not desire to eat the pineapple. Now, let us suppose further, that if she had tasted pineapple, she would prefer it to watermelon. It seems that a theory of well-being should say that her well-being would improve more if she were to eat the pineapple than if she were to eat the watermelon.

Actual intrinsic desire views have trouble explaining the hypothetical case under discussion, because the agent has no desire for the pineapple. Hence, her well-being, it seems, can only increase if she eats the watermelon. Perhaps a proponent of an actual intrinsic desire view would respond that what accounts for our intuitions in cases such as the one just discussed is that there is some other intrinsic desire that we are presupposing the agent has. For example, we might be presupposing that the agent has a desire for pleasurable experiences and that, were she to taste the pineapple, she would have a

pleasurable experience. So, the reply by the proponent of an actual intrinsic desire view goes, our intuitions about well-being do not conflict with an actual intrinsic desire view in this case because we can presuppose intrinsic desires that everyone shares in our thinking about the “pineapple” case.⁴

The reply under discussion is clever, but unfortunately it will not work in every example. To show that it will not work, consider a case of a desire for a life-long project where there is not another desire that one can point to in order to explain our intuitions in the case. Take for example, a librarian who does not prefer to be a lawyer even though he would if he knew more about how his life would go if he were a lawyer. In this case, there is no plausible alternative desire that one can point to, in all of the various more fully-detailed explanations of the example, in order to explain the intuition that the man’s life would go better if he were to be a lawyer. So even though the librarian does not desire to be a lawyer, the intuition goes, his life would go better if he were to be a lawyer. Later, I will discuss what I take to be driving our intuitions in the cases discussed in this part. But for now, it is enough to be clear that there is something wrong with actual intrinsic desire views.

I think there is another problem with actual intrinsic desire views. I will draw from Richard Brandt’s book, *A Theory of the Good and the Right*, where he argues one’s intrinsic desires can be lost if one were to be informed with propositional knowledge.⁵

Brandt thinks that if someone were to undergo cognitive psychotherapy then it would be

⁴ I do not mean to commit myself to the claim that absolutely everyone shares the desire for pleasurable experiences. Certainly people might not share such a desire. An ascetic, roughly, is someone who renounces material comforts. An ascetic, it seems, might not desire pleasurable experiences. For simplicity, I have granted the universal holding of the desire.

⁵ Although many parts of the book touch on this subject, Chapter VI “The Criticism of Pleasures and Intrinsic Desires” is most apt.

possible for some of his intrinsic desires to change. The process of cognitive psychotherapy is described by Brandt in the following quotation.

[The] aim is to show that some intrinsic desires and aversions would be present in some persons if relevant available information registered fully, that is, if the persons repeatedly represented to themselves, in an ideally vivid way, and at an appropriate time, the available information which is relevant in the sense that it would make a difference to desires and aversions if they thought of it. By 'ideally vivid way' I mean that the person gets the information at the focus of attention, with maximal vividness and detail, and with no hesitation or doubt about its truth. I mean by 'available information' the beliefs associated with this term in Chapter 1: relevant beliefs which are a part of the 'scientific knowledge' of the day, or which are justified on the basis of publicly available evidence in accordance with the canons of inductive or deductive logic, or justified on the basis of evidence which could now be obtained by procedures known to science. (Brandt, pg 111-112)

Elsewhere, Brandt describes the process of cognitive psychotherapy in more detail, but this quotation should provide us with what we need for now. Brandt is suggesting that intrinsic desires can be lost, in principle, by undergoing a process of being informed with all relevant available information. This could give us a counterfactual account of the relevant desire set. For our purposes, it does not really matter whether Brandt is concerned with a theory of well-being. For what it is worth, in the passage quoted immediately above, Brandt is only concerned with an account of rational desires and he is not directly concerned with a theory of well-being. The counterfactual account of well-being would be, roughly, that any and all desires that *would* survive cognitive psychotherapy are in the relevant set. Note that information used in cognitive psychotherapy is heavily restricted on Brandt's description. I will return to the issue of the nature of the information someone is to be subjected to in the process of cognitive psychotherapy in 3.2. For now, I wish to focus on the sorts of cases Brandt has in mind in the quotation above. The process of cognitive psychotherapy should be a bit clearer after a detailed discussion of some hypothetical cases.

Brandt gives four cases where he thinks that desires can be criticized as irrational, in that they might not survive cognitive psychotherapy or at least might not remain the same after cognitive psychotherapy. I will look at two of those four cases. Although Brandt's chapter is geared towards criticism of intrinsic desires and pleasures, some of his examples look like examples of instrumental desires (and pleasures). I will have to do some explaining in these cases. It should be noted that Brandt takes his four examples to be neither mutually exclusive nor exhaustive of all of the ways of criticizing intrinsic desires.

The first of the two cases I will examine involves a dependence on false beliefs (Brandt, pg. 115). Consider a hypothetical example of someone who pursues a Ph.D. in an academic field because he thinks (falsely) that his parents want him to do it; now, as I will explain, he wants an academic life intrinsically (Ibid, pg. 115). Brandt thinks that if this person were to be subjected to information that his parents would not have been disappointed had he pursued something else, then he could cease to desire to get his Ph.D.⁶

Brandt's example *might* look like a simple case where he is merely talking about instrumental desires. But let us suppose that it is not. The agent's belief that his parents want him to pursue a Ph.D., and his desire to please his parents, are perhaps crucial in his formation of a desire to pursue a Ph.D. We can treat that belief and desire as *explaining* how he came to have the desire to pursue a PhD without interpreting that desire as instrumental. Similarly, someone might have an intrinsic desire that is caused by a

⁶ Other similar examples from Brandt include a case of a person who develops a dislike of a particular food because eating it once made him ill, and a case where one has a hard time enjoying himself because he believes God wants him to live an ascetic life.

chemical imbalance. This chemical imbalance might explain, from the standpoint of an observer, how an agent would come to have this desire without the chemical imbalance being the ultimate end of the desire. In Brandt's example, the desire to please his parents might be the cause of his current desire to get a Ph.D. But if the agent, on Brandt's account, would desire to get his Ph.D. in the appropriate circumstances even if he did not believe that satisfying his desire to get his Ph.D. would lead to the satisfaction of some other desire, it is intrinsic.

Brandt does not claim, in the above described case, that such a person, upon confrontation with full available information, must absolutely give up his intrinsic desire to pursue a Ph.D. Brandt is merely claiming that it is possible that he would. The extent to which people abandon intrinsic desires such as these is a separate issue from whether abandoning these kinds of desires is possible. So Brandt's first case seems to be a very clear-cut case of how someone could cease to desire something intrinsically. Let us look at one more case.

The second case I will look at is Brandt's fourth case. So I have skipped over his second and third cases. This fourth case is based on exaggerated valences produced by early deprivation (Ibid, pg. 122). According to Brandt, one may acquire intrinsic desires that are stronger than they otherwise would be, when sufficient deprivation of the desired object occurs early in one's life. For example, someone who grows up poor may value wealth disproportionately to the way she would have if she had not grown up so poor. Further information, perhaps even knowledge of the fact that she would value money differently if she had not grown up poor, could make her cease to value money as much as she does. In this particular example, the person who what a poor child does not reject

of an intrinsic desire, but the example is sufficient to show that the hierarchy of a person's desires could change with full relevant available information. Brandt does, however, discuss a case in which, over time, he suspects one could come to reject an intrinsic desire that was formed from deprivation (Ibid, pg. 124). However, the example at hand involves a mere change in intensity.

At this point there are two general classes of desire-satisfaction theories under consideration: a class in which one's actual intrinsic desires constitute the relevant desire set and a class in which desires one *would* have under full information constitute the relevant desire set. I will argue that the second, counterfactually construed, class is more plausible.

In light of the example above involving the librarian, actual intrinsic desire views seem incomplete. I think that an appeal to the concept of *personality* should help make things clearer. A personality, roughly, is a set of dispositions. I think that someone's personality, together with her set of experiences and beliefs, generates her desire set. This is *close* to an idea Railton advances toward the end of his "Facts and Values." Railton says: "[A personality] is a collection of properties that ground dispositions to react in various ways to exposure to certain facts" (Railton, "Facts and Values," pg. 60). Railton sees a personality as that which *grounds* dispositions to react in certain ways under certain conditions. I see a personality the set of dispositions itself. I will return to Railton's account of personality in 4.6. He and I differ on some substantive issues about the role of the concept of personality in desire-satisfaction, specifically ideal advisor, theories of well-being. However, I suspect that Railton would find my development of

the concept of personality and my account of its role in desire-satisfaction theories friendly.

I agree with Brandt that one's intrinsic desire set can change given a change in one's information. Something must be behind the change from having a desire to not having a desire, given a change in information. We do not want to say that, a change in a person's information can, by itself, change his or her desire set. There must be something else. I say what else is needed for an explanation of change in one's desire set upon receiving more information is the concept of a personality. One's personality is what forms desires given beliefs and experiences. I think that actual intrinsic desire-satisfaction views can fail to reflect one's true personality.

Let me explain. In a way, actual intrinsic desire views do not capture the full range of one's *personality* in the way that a counterfactual theory captures the full range more easily. It goes without saying that people have a limited range of experiences and information. People's personalities are often compatible with a much larger range of life plans or projects than one actually explores and adopts. I think one underlying idea for desire-satisfaction accounts of well-being is that one's personality is an important factor in how well one's life is going. An actual intrinsic desire set is incomplete in a way: it does not reflect the full range of the personality. I think that the example of the librarian who would prefer life as a lawyer, were he to know more his potential life as a lawyer, nicely illustrates the incompleteness of actual intrinsic desire views. The person's life would go better if she were a lawyer, I say. Being a lawyer "fits" with her personality. The counterfactual account of the relevant desire set can capture the full range of one's personality and bring it to bear on one's well-being.

I think that Brandt's examples of the academic and person with an impoverished childhood illustrates a different, but related, problem with actual intrinsic desire views. As I say above, one's personality together with his or her experiences and beliefs will generate a desire set. Now, different desire sets could result when one's personality is combined with different belief and experience sets. Not all desire sets will reflect the true personality, I claim. Many belief or experience sets that include false beliefs or inadequately varied experiences will *distort* one's personality. It is only the desires that are true to one's personality that should be relevant to one's well-being: only the satisfaction of desires that are true to one's personality should count as increasing one's well-being.

My goal of 3.1 is to criticize actual intrinsic desire views. Theories that treat the desires one would have under certain counterfactual conditions do much better in two ways. The desire set one has in certain counterfactual conditions better captures the full range of an agent's personality than do actual desire views. Additionally, the desires one would have under some, but not all, counterfactual conditions reflect one's true personality better than do actual intrinsic desire views: there is less distortion on some counterfactual theories.

Now, I understand that I have not said much about what constitutes a true or genuine personality. At this point, I have only relied on a rough notion of personality—personality as consisting in a complex set of characteristics or dispositions that make up a person. Of course, I owe a more rigorous account of what a personality consists in. A more rigorous account will be developed in the next chapter.⁷

⁷ I take myself in Chapter 2 of this dissertation to have shown that desire-satisfaction accounts of well-being are superior to mental state accounts, explanatory objective accounts in the form of human nature

In this part, I hope to have made clear that the satisfaction of extrinsic desire, as such, do not increase well-being. Only the satisfaction of intrinsic desires increases well-being. Also, I hope to have shown why counterfactual accounts of the relevant desire set are superior to accounts that identify the set as composed of actual desires. I think that lives can go well in ways that agents do not actual desire and I hope to have made clear that the satisfaction of at least some actual desires does not increase well-being. Brandt's theory that the desires one would have under certain counterfactual conditions is much better at capturing our intuitions in many cases. The explanation of our intuitions in such cases is that one's personality is central to one's well-being. There are other counterfactual accounts besides Brandt's. I critically compare and contrast several sorts of counterfactual accounts in 3.2.

3.2 Counterfactual Desire Accounts Developed

3.2.1 Information and the Relevant Desire Set

With 3.2.1, I begin to examine counterfactual accounts of the relevant desire set. There have been a number of counterfactual accounts proposed in recent years. I will begin with a critical discussion of aspects of Brandt's version as a starting point and then move into more recently developed theories. I remind the reader that Brandt, in *A Theory of the Good and the Right*, does not endorse a desire-satisfaction account of well-being. However, what he says in his account of rational desires can be adapted to an account of

accounts, and pluralist accounts. The reader who finds my arguments in Chapter 2 inadequate might worry about what is driving the arguments to my preferred desire-satisfaction theory. The skeptical reader might worry that some other conception of well-being is really the driving force behind the arguments. As my favored version of desire-satisfaction accounts gets more sophisticated, it is important to bear in mind that I think that the personality is one central driving force. The path from thinking of the relevant desire set as actual intrinsic desires to thinking of the relevant set as somehow counterfactual is indeed complicated, but I think that the idea that personality is relevant to well-being helps make it clear. I will return to this general topic again.

well-being. So the counterfactual theory of well-being under discussion is merely inspired by what Brandt says.

I will now take up one aspect of Brandt's idea of "all available information"—his idea that one undergoing the process of cognitive psychotherapy be subjected only to "the propositions accepted by the science of the agent's day, plus factual propositions justified by publicly accessible evidence (including testimony of others about themselves) and the principles of logic" (Brandt, pg. 13). A plausible alternative to Brandt's account of "all available information" would be to include all true propositions and perhaps even knowledge of what things are like. We can divide knowledge into two camps: propositional knowledge (knowledge that . . .) and non-propositional knowledge (knowledge how . . ., or knowledge what . . .). Examples of this distinction seem intuitively plausible. For example, we can know *that* X is a piece of pineapple, and this is distinct from knowing *what* pineapple tastes like. Given the quotation above, I am very much *inclined* to read Brandt as saying that all of the information to which one is to be subject is propositional. Perhaps I am mistaken in my interpretation of Brandt and if so there is not a major problem here. I will discuss the prospects of including non-propositional information in the next chapter. Presently I will deal only with propositional beliefs. The task at hand then is to explore the prospects of Brandt's restriction of the information set to that which is supported by the best science of the day, publicly accessible evidence, and principles of logic.⁸

Brandt's reasoning on this issue is a bit unclear. Here is what he says about it:

⁸ I will follow Brandt in using the term "information" very loosely in that it allows false propositions to be information. Strictly speaking, this usage is probably incorrect, but it is a simple way of speaking.

It [all available information] is also contrasted with the beliefs an omniscient being might have, beliefs in all true propositions, since we do not know the identity of such beliefs and can hardly use them as a tool of criticism. (Almost everything a person does would be different if he had the knowledge of an omniscient being made available to him—for instance, if he were informed of the cure for cancer.) We might however, identify all available information with the factual beliefs an agent would have had at the time of action, if his beliefs had been fixed by his total observational evidence and the principles of logic, both inductive and deductive. It seems more useful, however, to define the concept to include beliefs the agent would have if he obtained evidence which he could obtain at the time. So I prefer to define ‘all available information’ as the propositions accepted by the science of the agent’s day, plus factual propositions justified by publicly accessible evidence (including testimony of others about themselves) and the principles of logic. (Ibid, pg. 12-13)

Brandt admits that this final definition of “all available information” does not exclude false propositions as being in the mix (Ibid, pg. 13). The science of the day may very well be seriously mistaken on many issues, so false propositions are bound to creep in to the information with which one is to confront her desires. That fact should be disconcerting, but there are other problems as well.

Brandt argues that we ought not to adopt the view that includes all true propositions in the information set. His reasoning seems to be that we cannot know what these propositions are and cannot use them as a tool for criticism. I think Brandt is mistaken in thinking that we need to give an account that we can use as a tool for criticism. Let me explain.

Brandt’s language in the above quotation suggests that he is after something we might call a “decision procedure” for finding out content of the relevant desire set. A description of a decision procedure, on my construal, would involve a description of an actual process one could go through in figuring out the relevant desire set. A decision procedure for identifying the relevant desire set is distinct from what I will call a “criterion” of the relevant desire set. A criterion for the relevant desire set is a criterion

that a desire has to meet in order for it to be a member of the relevant desire set. One helpful way of understanding the distinction between a decision procedure and a criterion is that a decision procedure is used to “track” something whereas a criterion simply “makes” something the case.⁹

When Brandt writes of the need to identify the information that the patient is subjected to in the process of cognitive psychotherapy, and the need for a tool of criticism, it seems that Brandt is suggesting that we need a decision procedure, or perhaps both a decision procedure and a criterion. It is hard to tell what Brandt is after based on what he says.¹⁰ Perhaps Brandt and I are after different things with our analyses, but I am not after a tool of practical criticism. It would be nice to have a tool of practical criticism, but developing such a tool is beyond the scope of this dissertation.

Neither Brandt’s model nor a full true propositional model could serve as a decision procedure for the relevant desire set. Both Brandt’s model and a full true propositional model include far too much information for one to actually be subjected to. This should not prove troubling. At this phase of the dissertation, I am after the best account of the relevant desire set. In order to provide the best account of the relevant desire set, I must provide a criterion any desire must meet to be in the relevant set. As I said, I am not entirely certain as to what Brandt is after, but it is clear that he cannot be providing a plausible decision procedure someone could go through in figuring out what

⁹ My distinction between a decision procedure for the relevant desire set and a criterion of the relevant desire set is modeled on a similar distinction in the debates about Utilitarianism. Some have felt that Utilitarianism is implausible because no one could know all of the relevant consequences of an act, thus could never figure out what to do. Many Utilitarians have replied that the Utilitarian principle is a criterion of right acts and not a decision procedure. The decision procedure, so the reply goes, is something to be worked out at a later time.

¹⁰ Brandt does admit that some of the information would be “intolerably expensive to get” (Ibid, pg. 13).

one would desire after undergoing the process of psychotherapy. Clearly, the set of information Brandt has in mind is far too large for anyone to actually process in real life.

Let us just assume that Brandt is proposing a criterion for the relevant desire set, and not a decision procedure. There still remains the issue of whether Brandt's construal of the information set is better than a full propositional information view. I think that a full propositional information view is far superior. Brandt allows a person's desire set to be influenced by false information, provided that the false information is countenanced by the best science of the day. While the reader might feel that much of the information of the best science of *today* is quite good, I ask the reader to consider someone undergoing Brandt's process of cognitive psychotherapy based on the information countenanced by the best science of one thousand years ago.¹¹ The patient would be subjected to all sorts of false propositions and form desires for things (or cease to desire things) in light of these false propositions. Additionally, Brandt's construal of the information set still might not be enough information. Let me explain.

As I say in 3.1, one's personality is central to her well-being. Actual intrinsic desire views are problematic because the desire set they generate might not capture the full range of one's personality and the desire set might distort one's personality. Similar problems arise for Brandt's account.¹² The set of propositions countenanced by the best science of the day, propositions justified by publicly accessible evidence and principles of logic, at least in many cases, will not capture the full range of agents' personalities and

¹¹ Indeed, if we think about the substantial span of human history when there was no science—at least no rigorous scientific method, Brandt's theory seems problematic simply because it is not clear what his theory would say about someone's well-being who lived in that time.

¹² Please bear in mind that the theory under criticism is inspired by Brandt and is not one that Brandt finally endorses.

might even distort their personalities. Forming desires (or ceasing to desire things) based on false propositions could lead a patient to desires that are not true to her personality.

As I have suggested, it might help to think of a personality as a set of dispositions or characteristics. If one of the core ideas behind desire-satisfaction accounts of well-being is that one's life goes better to the extent to which the world and the desires stemming from one's personality fit, then Brandt's theory is problematic in a similar way in which actual intrinsic desire views are problematic: the desire sets of both theories are often either not representative of the full range of the personality or desires in the set distort the personality.

So, the relevant information set should be full information as opposed to the best information available given the science of the day (plus propositions justified by publicly accessible evidence and principles of logic). The account of well-being that treats one's well-being to vary to the extent to which the desires one would have after having full information are satisfied is sometimes called an "ideal observer" account of well-being. Put roughly, ideal observer accounts of well-being hold that one's life goes well to the extent to which the desires one would have if fully informed are satisfied. There are several aspects of Brandt's account that remain to be critically discussed. First, as of yet, there has been no serious critical discussion of the nature of the process by which one is to be fully informed. Brandt has in mind a process of cognitive psychotherapy that involved repeated exposure to ideally vivid information (Ibid, pg. 111-112). Brandt's claim is that the relevant desire set is constituted by desires one would have after undergoing such a process. This claim is extremely controversial. I will critically discuss Brandt's explanation of the process and several recent criticisms in the next

chapter. Second, there has been no significant discussion of whether to include non-propositional knowledge in the information set. Thus far, the discussion has only been about what sort of propositional information to include. I will deal with the matter of non-propositional knowledge in the next chapter. What I will turn to next is the distinction between “ideal observer” views and “ideal advisor” views. I will argue that ideal advisor views are superior to ideal observer views.

3.2.2 Ideal Advisor Views

Ideal observer views hold that one’s life goes well to the extent to which the desires one would have if fully informed are satisfied. Ideal observer views are far more plausible than actual desire views, but that does not mean that ideal observer views do not have problems of their own. I prefer a different sort of account—ideal advisor views of the kind proposed by Peter Railton. Railton characterizes ideal advisor views, roughly, as those according to which the relevant desire set is constituted by the desires one would want herself to have, as she is in her actual circumstances, were she to have full information. Ideal advisor views get quite complicated. I would like to motivate the need for such view by pointing out problems with ideal observer views.

The desire set of the ideal observer and the set “recommended” by the ideal advisor would likely differ, and for any individual, they might differ sharply. The following is an example to illustrate this. The following example is adapted from Alan Gibbard in his *Wise Choices, Apt Feelings*. Imagine that a person desires to eat. Some of the information her fully informed counterpart will have is knowledge, in maximally vivid detail, of the inner workings of the digestive system. Her fully informed counterpart may

find this disgusting and cease to desire to eat.¹³ However, presumably her fully informed counterpart will want her less informed, actual self to want to eat. Thus, taking the advice of the ideal advisor as constituting the relevant desire set for well-being can get around certain problems with the ideal observer account. I do not intend the example I have adapted from Gibbard to conclusively refute ideal observer theories, but it does seem to cause a problem for the theory.

Another problem for ideal observer theories comes from Peter Railton in his “Facts and Values” (Railton, “Moral Realism,” pg. 36). Railton notes that an ideal observer, being fully informed, will have no desire for more information. Here is what Railton says: “For example, [an ideal observer] presumably does not want any more information for himself—there is no more to be had and he knows this. Yet it might still be true that [an ideal observer] would want to want more knowledge were he to be put in the place of his less well-informed self, [his actual self]” (Railton, “Moral Realism,” pg. 36). Let me clarify. It is important to distinguish a desire for information from a desire for *more* information. Only the latter, I say, causes problems for ideal observer theories. An ideal observer can desire information, even as he is—that is to say, fully informed. A person can desire at t_1 a state of affairs that he knows obtains at t_1 . We might think of a person’s desire for something he already has, and knows he has, as an instance of “cherishing,” or some other similar state of mind. Nevertheless, given that I construe desires quite broadly, “cherishing” something is a way of desiring that thing. A desire that one have *more* information is quite different, however. An ideal observer’s desiring more

¹³ Gibbard’s example, from page 19 of his book, has more detail and complexity than I have presented. I have simplified the example for convenience. Also, Gibbard construes his example as going against Brandt’s characterization of “rational desires” and is not specifically formulated to go against ideal observer theories. However, I have adapted it for my purposes.

information is a desire of that which is impossible and that which he or she knows is impossible. I do not know whether it is possible to desire that which one knows to be impossible—or, worse, logically impossible. If it is indeed impossible to desire that which one knows to be impossible, then ideal observer theories exclude some desires, simply in virtue of the process of being fully informed. Even if it is possible to desire that which one knows to be impossible—so, some ideal observers could desire to have more information, a problem remains for ideal observer theories. An ideal observer may lose his or her desire for new information upon being fully informed: that is still a problem. It seems an implausible feature of ideal observer theories that they allow for an ideal advisor to lose his desire for more information. An actual agent's desire for more information, on ideal observer accounts, whether would not or might not be relevant to his or her well-being. That is an unattractive feature of ideal observer accounts.

3.2.3 What the Ideal Advisor Advises (or what advice we should listen to)

Ideal advisor views can be fleshed out in different ways. Peter Railton, in his “Moral Realism” and part of “Facts and Values” talks about desires an ideal advisor wants his actual counterpart to have (i.e. which desires the ideal advisor desires for you). Let us call this specification of the relevant advice the “desire to desire” account. If we treat the desire to desire account as holding that well-being is increased when the ideal advisor's desire is satisfied, there are significant difficulties. What happens when an ideal advisor's desire is satisfied is that his or her advisee merely forms the desire that the advisor wants the advisee to have. For example, imagine that my ideal advisor would desire that I desire that I take a swim. When the ideal advisor's desire is satisfied, I desire to take a swim. Presumably though, and this is the intuition here, my life goes

better in this case where I take a swim. Well-being could never depend *entirely* on forming and losing desires.

An additional bit of clarification is in order. My ideal advisor might desire for instrumental reasons that I desire to swim—his reason might be that when I desire something, at least some times, I bring it about. The satisfaction of such an instrumental desire cannot enhance one's well-being because there is the problem of double counting mentioned earlier in this chapter. Desires that are in the relevant set must be intrinsic in some way.

Here is one more clarification about the “desire to desire” understanding of an ideal advisor view. Let us say that *A* is an actual person and *A+* is his fully informed idealized self. There could be cases where *A+* desires something for *A*, but *A+* does not desire *A* to desire it. For example, Frankenstein's monster desires friends and has none. This makes him miserable. The monster's ideal advisor might very well want him to have friends, but not want him to want to have friends. Although these sorts of cases must be stated in a fairly complex way, the intuition seems fairly clear and straightforward. There are states of affairs such that desiring them makes one worse off, but having them makes one better off. The issues revolve around how unlikely the state of affairs is to obtain and how miserable the knowledge that one cannot satisfy the desire would make the agent. There is a great deal of difficulty in trying to capture the advice of the ideal advisor in the type of case found in the Frankenstein's monster example when using the “desire to desire” language.

Given all of this, it seems that on the “desire to desire” interpretation, ideal advisor views are implausible. I think that the philosophers who write on ideal advisor views

must be up to something else. Perhaps a better way of talking about ideal advisor views is to drop the desire to desire language altogether.

Here is a more charitable way of understanding what ideal advisor views are about. Instead of taking the state of affairs in which the ideal advisor's desire is satisfied as determining the well-being of the advisee, take the object of the advisee's desire. Look at the example again: *A*+ desires that *A* desire to take a swim. Instead of taking *A*+’s desire here, take the desire *A* would have if *A*+’s desire were satisfied (call this the “second desire”). The second desire is satisfied when *A* takes a swim. The intuitive idea is that one’s well-being is increased when the state of affairs expressed by the object proposition of the second desire obtains. Let me clarify. In our examples, *A*+ has a desire but the satisfaction of her desire is irrelevant, at least in many cases, to *A*’s well-being. The more intuitive way of understanding ideal advisor views is to treat the object of the second desire as important. Railton and others who write on ideal advisor views must have in mind this second interpretation of ideal advisor accounts.

There is a way to simplify all of this ideal advisor talk. We can take out the problematic and confusing element. Instead of talking about what *A*+ desires that *A* desires, we could instead talk about what kinds of things *A*+ wants for *A*.¹⁴ The clearest way of talking about ideal advisor accounts is to say that *A*’s well-being varies with the extent to which *A* lives in accordance with what *A*+ desires intrinsically for *A*. Call this the “streamlined” take on ideal advisor views. The result would still be a type of desire-satisfaction account and it takes (at least some of) the desires of *A*+ as determining what constitutes *A*’s well-being.

¹⁴ Sobel mentions this in footnote 19 of his “Full Information Accounts of Well-Being,” as does Rosati in her “Internalism and the Good for a Person.”

Let me return to a few examples discussed above and show how the “streamlined” interpretation of ideal advisor views deals with them. First, consider the examples involving reading the newspaper and eating one’s food. The streamlined account explains them very well. $A+$ will not need to read the newspaper and he might lose his appetite, but if $A+$ would still want A to read the newspaper and eat, then it is plausible to think that those two states of affairs are relevant to A ’s well-being. Second, if we look at the Frankenstein example, Frankenstein’s ideal advisor wants him to have friends but does not want him to want to have friends. In such a case, Frankenstein’s having friends will increase his well-being, but wanting them, as such, without getting them, will not. Third, if $A+$ wants A to take a swim, then A ’s taking a swim would increase A ’s well-being, but merely wanting to take a swim, as such, would not.

One further benefit of a streamlined ideal advisor view is that it coheres well with our general intuitions about personality. $A+$ is A with additional information. Someone’s idealized counterpart has far greater knowledge than her partially informed self. Despite this great difference, a person and her idealized counterpart have a lot in common as well. They share a personality, I say. If someone and her fully informed counterpart share a personality, then the fully informed person is ideally situated to speak on matters of her partially informed counterpart’s well-being.

Now, what I say about personality and well-being is contentious and will have to be defended in the next chapter. Ideal advisor views have been criticized recently in the literature. Some objections turn on the possibility of any sort of process by which one could become fully informed. Some objections turn on the very intelligibility of ideal advisor views. There have been some objections that turn on whether $A+$ ’s position is

authoritative for *A*'s well-being. Finally, there have been some objections to the effect that ideal advisor views do not remain true to one core motivation for desire-satisfaction views more generally. I will address all four of these objections in the next chapter.

CHAPTER 4 IDEAL ADVISOR VIEWS

In Chapter 3, I looked at several different kinds of desire-satisfaction accounts of well-being and concluded that an ideal advisor account is superior to rival desire-satisfaction views. In this chapter, I will examine several recent criticisms of ideal advisor views. I will argue that the account developed in Chapter 3 either is not susceptible to the criticisms offered or that a proponent of the account developed in Chapter 3 has responses to the criticisms available to him.

4.1 The Internalism Requirement

Connie Rosati, in her paper “Internalism and the Good for a Person” and Don Loeb in his paper “Full-Information Theories of the Individual Good” discuss an “internalism requirement” that they feel any adequate account of what is good for a person must meet. Most, if not all, non-desire-satisfaction conceptions of well-being would be eliminated in virtue of the fact that they do not meet the internalism requirement. Only specific kinds of desire-satisfaction accounts meet the internalism requirement, according to Rosati and Loeb. Specifically, they claim that no full information ideal advisor view meets the internalism requirement. If they are right, then much of my work in Chapter 3 of this dissertation has been a waste of time.¹

¹ Interestingly, if there were to be an internalism requirement of the sort that Rosati and Loeb endorse, then much of my work in Chapter 2 could be correct in its conclusion but needlessly complex in its arguments. My work in Chapter 2 would have been correct in its conclusion because I would be right in thinking that some desire-satisfaction accounts of well-being are superior to non-desire-satisfaction accounts of well-being. Following the line of thought, my work in Chapter 2 would have been needlessly complex because so many theories of well-being would be excluded simply in virtue of the fact that they fail to meet the internalism requirement.

Very roughly, an internalism requirement would make a strong or essential connection between the good for a person and the person's motivation. Both Rosati and Loeb provide brief introductory definitions of the internalism requirement they have in mind. Rosati goes on to provide a more precise definition and Loeb does not. Before I go into Rosati's and Loeb's arguments, I would like to explore some comments by Peter Railton. Railton's work has influenced Rosati's and Loeb's work on the internalism requirement.

Both Rosati and Loeb are heavily influenced by Peter Railton's arguments in his papers "Facts and Values" and "Moral Realism." Here is the quotation from Railton's "Facts and Values" that Rosati gives and Loeb refers to:

It does seem to me [Railton] to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him. (Railton, "Facts and Values", pg. 47)

Railton does not intend to give a precise account of an internalism requirement. He does hint that there must be a connection between motivation and what is good for a person. The terms "compelling," "attractive," and "engage" are a bit vague and perhaps he intends to leave the account open ended.²

Loeb thinks that the internalism requirement is that an adequate theory must show there to be a strong connection between what is good for a person and his motivation.

² The objective list theorist could argue that if people are rational and aware, in some sense, of their goods (such as wisdom and friendship), then they would be engaged, at least in some sense, by them. The proponent of an objective list theory probably cannot show a necessary connection between a good and a motivation for that good, but, Railton's language is weak here and strictly speaking would allow room for such a view. So, Railton should be understood in a bit narrower sense than the broadest sense possible. Please see Rosati's take on a more precise account of the connection between a person's good and his motivation below.

Loeb characterizes a requirement as follows: “The first [attractive feature of full information desire-satisfaction theories] is motivational. The theories typically embrace a version of internalism. They characterize a person’s good (or reasons for action, etc.) in terms of her motivation. The motivation in question need not be actual. Dispositions to be motivated would suffice” (Loeb, pg. 1). Loeb clearly has in mind a strong internalism requirement. It is not clear exactly what he has in mind when he says “They characterize a person’s good . . . in terms of her motivation.” Perhaps all that Loeb has in mind is that if something is good for someone (A), then A has some motivation to pursue it. But Loeb might have something stronger in mind. Rosati construes the internalism requirement in a way similar to Loeb’s, but in clearer way.

The internalism requirement, on Rosati’s account, is “the general thesis that there is a necessary connection between motivation and normative status. The thesis tells us that X can have a certain normative status N only if someone A would be motivated by it in sense M” (Rosati, pg. 298).³ Let us follow Rosati in calling this kind of internalism “existence internalism.” This terminology comes from Stephen Darwall in his “Reasons, Motives and the Demands of Morality.” He defines “existence internalism” roughly as the position that there is a necessary connection between the truth of a normative proposition and motivation (Darwall, pg. 308). Darwall divides existence internalism into two groups: *perceptual internalism*, according to which “it is impossible for a person to know directly or perceive the truth of a normative proposition without being moved” and *metaphysical internalism*, according to which “an agent’s being moved (under appropriate conditions) is either part of what it is for a normative proposition to be true of

³ When I use direct quotation, I will leave “A” as it appears in the original, rather than add italics as I do for ease of reading.

him, or a necessary condition for the holding of that normative fact” (Ibid, both page 308). I suspect that Rosati has in mind a sort of metaphysical internalism, but it is not absolutely clear.

The existence internalism requirement, on one rough formulation, is that something can be good for *A* only if *A* can be motivated to pursue it. It *seems* that the internalism requirement, even in its rough formulation, would cause problems for ideal advisor views such as mine and Railton’s.⁴ As I suggest in the preceding chapter of this dissertation, I think that intrinsic desires can be lost and gained through a process of being fully informed. As a result of this changing of one’s intrinsic desire set, an ideal advisor can have desires that her advisee does not have. Reverting back to the language of the preceding chapter, I say that *A*’s advice might not overlap with *A*’s desires. Moreover, *A* can want something for *A* without wanting it for herself (*A*!). So it might seem that, if an internalism requirement of the sort that Rosati and Loeb thinks exists does indeed exist, then ideal advisor theories have some explaining to do.

The internalism requirement, very roughly, links one’s good with one’s motivations. The formulation thus far explained, is still too vague for our purposes here. Let me identify two different versions of the internalism requirement: the “weak” version and the “strong” version. The weak version of the requirement is that in order for something to be good for someone, it must be possible, in some way, for her to care for it. The weak version, I say, is that it must be *logically possible* for one to care about something for it to be relevant to his well-being. Rosati offers a strong version of the internalism requirement. I call Rosati’s position “stronger” because it treats the

⁴ For what it is worth, the internalism requirement seems to cause similar problems for ideal observer views as well.

connection between one's good and one's motivations in a more specific, narrower, way than mere logical possibility. She thinks that there are two internalist conditions an account of the personal good must meet.

Rosati has in mind that X is good for a person A only if,

1. were A under conditions C and contemplating the circumstances of her actual self as someone about to assume her actual self's position, A would care about X for her actual self; and
2. conditions C are such that the facts about what A would care about for her actual self while under C are something A would care about when under ordinary optimal conditions. (Rosati, pg. 307)

Rosati calls her version of the existence internalism requirement "two-tiered" because it has two clauses. Rosati explains "ordinary optimal conditions" as conditions in which an agent is thinking rationally, calmly and not ignoring or overlooking relevant information (Ibid, pg. 303). Ordinary optimal conditions seem to be, roughly, conditions of clear, careful thinking.

To clarify Rosati's position, let us say that A_o is A in optimal conditions.⁵ Recall that optimal conditions are not the informational conditions. Let us say that on Rosati's account the advisor is A_{ic} , who is A in informational conditions and other counterfactual conditions, such as contemplating the circumstances of her actual self, as someone about to assume her actual self's position, etc. Now, Rosati requires that A_{ic} 's advice for A be something A_o cares about. On Rosati's account, if A_o does not care about what A_{ic} advises for A , then what A_{ic} advises cannot be good, as such, for A .

⁵ I prefer to have ' A_+ ' refer just to A 's fully informed counterpart. A_+ and A_o might differ significantly in several ways: perhaps most notably, A_o does not have to be as informed as is A_+ . A_o is merely thinking clearly and rationally, and not necessarily with full information.

Let me explain how Rosati's strong internalism seems to be incompatible with ideal advisor views such as Railton's and the one I develop in Chapter 3 of this dissertation. Traditional ideal advisor theories characterize a person, *A*, in two circumstances: *A* as he actually is and *A*+, that is, *A* as he would be with full information, deliberating without rational error, etc. Now, *A*_{ic} could be identical to *A*+, but nothing in ideal advisor theories, as they are traditionally understood, requires that *A* care about *A*+'s advice. So it looks like ideal advisor theories, as they are traditionally understood, do not necessarily meet Rosati's internalism requirement.

My ideal advisor account can accommodate the weak internalism requirement. If *A*'s having full information is possible, then, let us just say, there is a possible world in which *A*'s fully informed counterpart *A*+ exists. Suppose that *A*+ wants *x* for *A*. Then, according to the theory, *A*'s life goes better when she gets *x*. There is a strong sense in which *A*+ cares about *x*; after all, she recommends it above other alternatives. Since *A*+ is just *A* with full information, then in a very straightforward sense it is *possible* for *A* to care about *x* and she cares for it under certain non-actual conditions. Now, in this way, on my ideal advisor account, it will always be the case that *A could* care about *A*+'s advice: and so it follows on the full information theory that one *could* care about what is good for her. If all that the internalism requirement were to amount to is that one could care, in some such way, about what is good for him, then my ideal advisor view would meet the internalism requirement. Moreover, my ideal advisor account seems to meet the conditions of the internalism account of the exact wording of the quotations from Railton above. To repeat Railton's exact words on this: "It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage

him” (Railton, “Facts and Values,” pg. 9). Railton’s exact words commit him to nothing stronger than the weak internalism requirement and his and my ideal advisor accounts meet the requirement without problem. At this point, I will turn to Rosati’s arguments in favor of an internalism requirement.

4.2 Rosati’s Specific Internalism Requirement

Rosati offers five arguments for an internalism requirement. Sometimes it is not clear whether she is offering an argument for the weak or strong version, so I will try to take extra care in being clear on that issue. Most of her arguments are aimed at showing that there is an internalism requirement of the weak sort mentioned above—a version of existence internalism—rather than her own strong version of the requirement.

The first argument Rosati gives is an argument from “judgment internalism.” Rosati claims: “According to judgment internalism, it is a necessary condition on sincere judgment about a person’s good that the speaker normally have some inclination, not necessarily overriding, to promote or to care about that thing” (Rosati, pg. 310). Here is Rosati’s argument:

An account of the good for a person can succeed in explaining and preserving the inherent normative force of judgments about a person’s good only if it suitably constrains the possible objects of a person’s concern relevant to a determination of her good. To do so, it must link a person’s good to what she would care about, but not just under any conditions whatsoever. Rather, the conditions must be such that information about a person’s wants or reactions under them would matter to the actual person, at least if she were under ordinary optimal conditions (Ibid, pg. 311).

Rosati’s argument goes something like this: judgment internalism is a fact and that fact is best explained by existence internalism. Her argument is an inference to the best explanation.

Rosati is thinking about judgment internalism in an unconventional way. Judgment internalism is usually thought to be a thesis about judgments by *S* about *S*’s own good.

Rosati seems to make a much stronger claim: that a judgment by *S* about *P*'s good must motivate *S* at least to some extent. Rosati's version of judgment internalism is very controversial. I suspect that it is false, but I do not think that I need to provide a rigorous argument against it. She tries to offer an argument in favor of existence internalism and uses an unusual version of judgment internalism that is even more controversial to support it.

Can Rosati use a more plausible version of judgment internalism to support an internalism requirement? I suspect not. Let us look at two other versions of judgment internalism: (1) judgment internalism as it is traditionally understood and (2) a hybrid view between what Rosati says above and judgment internalism as it is traditionally understood. Let us start with (1).

Judgment internalism, as it is traditionally understood, is a thesis about a link between one's judgments of a certain sort and one's motivation. Darwall characterizes it as follows: "Judgment internalism holds that if *S* judges (or believes, or sincerely asserts) that she ought to do *A* (or that she has reason to do *A*), then, necessarily, she has some motivation to do *A*" (Darwall, pg. 308). Darwall does not mention judgments by *S* about what is good for *S* in this passage, and, of course, judgments by *S* about *S*'s own good are really the central sort of case for the present discussion.

Judgment internalism, as it is traditionally understood, does not support either the weak or Rosati's strong existence internalism requirement. People are often mistaken in their judgments about what is good for them. If judgment internalism were true, people would have to be motivated to pursue that which they mistakenly think is good for them. This suggests that judgment internalism is really about the nature of *judgments about*

what is good for someone, rather than what *really is good* for them. Recall that Rosati seems to think that the internalism requirement on what is good for someone is a kind of existence internalism, and she argues that existence internalism is the best explanation of judgment internalism. But this is a mistake. The best explanation of judgment internalism, as it is traditionally characterized, is not existence internalism—which is about what really is good. Instead, the best explanation of judgment internalism as it is traditionally understood must have its roots in the nature of judgments. So even if the traditional version of judgment internalism were to be true, it would not help Rosati prove that there is anything like an existence internalism requirement on what is good for someone.

Might there not be a version of judgment internalism that is more plausible than the version Rosati originally offers and yet at the same time could support an existence internalism requirement? It is hard to see how such a version could be formulated in a plausible way and still serve the required role in Rosati's argument. Rosati's version of judgment internalism could be made slightly closer to ordinary versions. She could say the following: All judgments by *A* in any possible condition (e.g. by *A*, *A*+, or *A* in some other possible circumstance) about the actual *A*'s good must motivate the judger (e.g. *A*, *A*+, etc.) to some extent. This hybrid version is closer to the traditional understanding of judgment internalism in that it limits the range of the "judgers" who must be motivated in different possible scenarios. The problem with this version is that it does not ensure *A*'s motivation in *A*'s actual circumstance. Alternatively, Rosati could say the following: All judgments by *A* in any possible condition (*A*, *A*+, or *A* in some other possible circumstance) about the actual *A*'s good must motivate the actual *A* to some extent. This

hybrid version will not help either because it is simply too controversial. A would not even be able to comprehend the full range of advice offered to him from those just like him in any possible condition. Recall that Rosati's argument is by an inference to best explanation and that she needs a theory that is at least fairly plausible to get her argument off the ground.

So Rosati's first argument to prove an existence internalism requirement does not work and cannot be fixed up in an easy way. Her original formulation of judgment internalism is very implausible. The traditional version of judgment internalism might not be right, and even if it is, it is not the case that the best explanation of it is existence internalism. Judgment internalism as it is traditionally understood and existence internalism on either the strong or weak version have quite different contents. There does not seem to be any obvious hybrid account of judgment internalism that helps Rosati's first argument.

Rosati's next argument is an argument from the metaphysics of value. The argument begins with the thought that value is subjective; that value is brought into the world with agents who value. Citing R.B. Perry's *General Theory of Value*, Rosati claims:

Introduce into the world creatures who are affected by and react to their world, Perry tells us, and you introduce value as well. Indeed, he suggests, introducing such creatures is sufficient for the introduction of value. It is a natural step from this thought to the thought that value itself must be a complex motivational property. (Rosati, pg. 313)

The only other possibility for the metaphysics of value that Rosati sees is Moore's non-natural properties. The problem with Moore's account of value is its queer metaphysical commitments. Rosati admits that the metaphysical argument is inconclusive because a proponent of dispositional accounts of value, such as McDowell and Wiggins, can

endorse the metaphysical view of value under discussion and yet still reject an internalism requirement of the good.⁶ A proponent of the dispositional account of value could agree with the metaphysical condition that agents bring value into the world by claiming that goodness consists in reactions or dispositions (such as one might claim that the color red is brought into the world by appropriately perceiving agents), without thinking that that kind of reaction or disposition is motivational.

I will not spend too much time on the metaphysical argument because Rosati admits it is inconclusive. She does offer some thoughts on how one might strengthen the argument, but her thoughts are not definite (Ibid, pg. 314-315). I think that we can agree with the metaphysical picture of value (namely, that introduction of creatures who are affected by and react to the world leads to the introduction of value), but disagree with existence internalism. Here is one thought that should make one think that this is a coherent position: we can distinguish between moral patients and moral agents. Moral patients, let us say, are those things that are deserving of moral treatment. Moral agents, let us say, are moral patients, but in addition to deserving moral treatment, they also can act morally and be held accountable for their actions. Only moral agents can be motivated (in the relevant sense). But moral patients react to and are affected by the world and presumably have a good. Thus value could be brought into the world by the existence of moral patients without there being the kind of motivation necessary for existence internalism. For if there were no agents, there would not be motivation in the right sense. It seems as though Rosati's view of the metaphysics of value implies that, if there were no agents, and at least some patients, there might not be any value. This

⁶ McDowell's and Wiggins's accounts are dispositional but not motivational

would be puzzling. I do not take myself to be proving that Perry's metaphysical picture of value is false. Merely that it does not support either weak or strong existence internalism.

Rosati's third argument is from the epistemology of value. We can justify the thought that something is good for someone, so the argument goes, only if the alleged good satisfies existence internalism (Ibid, pg. 316). Rosati's first of two lines of epistemological argument is inspired by Mill's supposed proof of the principle of utility, the argument that turns on the analogy between visible/seen, audible/heard, and desirable/desire. In Rosati's words:

Unless a person *could* care about the thing in question it cannot be justified as a part of her good, because the possibility of her caring about the thing is necessary evidence of its being good for her. But why think that it is necessary evidence? Consider the following thought experiment. Suppose that a person A could not be brought to care about a thing X under any conditions and so concluded that it is not good for her. What counterevidence could be produced to subvert A's conclusion? We have no picture, the argument might go, of what such evidence could be. (Ibid, pg. 316)

Rosati does not claim that the person must actually desire the thing in question; merely that it must be the case that it could be desired. The argument in the quotation seems fairly clearly to be an argument for weak internalism. However, I think it is, at best, far from conclusive. This is because epistemological arguments can only prove so much in this field. Humans might simply have a limited capacity for knowledge that does not extend to the full field of what exists. If someone cannot be convinced of the existence of something, then that does not mean that it does not exist.

However, suppose that the limits of human capacity for knowledge extend far enough to include all that is good for them. I still do not think that Rosati's arguments support a strong internalism requirement, and probably do not even support a weak

internalism requirement. I would like to distinguish between two ordinary uses of the term “care.” Sometimes when trying to convince people of something by showing evidence for it, people might respond “why should I care about that?” and what they mean is something like “why should I believe in that?” Imagine that the members of a jury are debating certain pieces of evidence that point to the guilt and innocence of the defendant. Someone may say “He stood to inherit a lot of money upon the death of his uncle” and another may respond “But why should we care about that.” It is in this epistemic sense that caring is more about belief than about motivation. Other times people speak of caring in a context where a conative attitude is meant, such as caring about whether cats are set on fire. If people could not be brought to see the importance of some evidence—if they could not be brought to “care”—in the epistemic sense, I would wonder whether it really is valid evidence. However if someone could not be brought to care and be motivated about something, that does not seem to be evidence that it is not a good—at the very least, additional argument is necessary.

Rosati’s second epistemological argument is inspired by Richard Brandt’s work on intrinsic desires and cognitive psychotherapy. Rosati states:

But Brandt’s strategy suggests that in order to justify to a person the claim that something is good for her one must be able to show how that thing connects with her own actual or possible concerns. We must show her that she has reason to care about it, and we can show her that she has reason to care about it only by appropriately connecting it with something that she already does or can care about. (Ibid, pg. 318)

We might view Rosati’s second epistemological argument as one that relies on *reasons* as opposed to what we can convince the person of. Now the concept of a reason, when left unexplained, will not help clarify the issue. For people say that there are many sorts of reasons.

Again, epistemological arguments can only show so much in this area and Rosati admits that epistemological arguments will be inconclusive (Rosati, pg. 318). However, I will try to deal with her argument directly. The plausibility of Rosati's argument depends on how we interpret "possible concerns" and on what it is to care about something. She does say this:

The reasons version of the epistemological argument yields a similar result. This argument relies upon the idea that we can show a person that she has a reason to care about X only by showing her that it connects with something she already does or at least could care about. Two-tier internalism exploits this very idea. If a person is to have a reason to care about X, and thus a reason to think that it is, at least *prima facie*, a part of her good, it is not enough that X be among the possible objects of her concern. Rather, it must be the case that although she does not now care about X, she would care about it under certain conditions, and what's more, the fact that she would care about it under those conditions is something that concerns her even now, at least if she reflects on the matter carefully" (Ibid, pg. 319).

In this passage, Rosati clearly makes a move to argue for the strong internalism requirement. However, there is very little of her argument that relies on an epistemological argument of any sort. Perhaps what Rosati is up to is proving a plausible way, at least by her lights, of fleshing out the details of internalism requirement. However this is quite different from providing an argument. So we must rely on her other arguments for a critical discussion.

Rosati's fourth argument is based on the "'ought' implies 'can'" doctrine. Drawing from David Velleman, Rosati writes:

We think of our good, he [Velleman] suggests, as being that which we ought, at least *prima facie*, to care about. Yet it cannot be that we ought to care about something if we are incapable of caring about it. We can be *prima facie* obligated to care about something only if it is at least *prima facie* an option . . . And something can only be *prima facie* an option for a person, if she is capable of caring about it. (Ibid, pg. 320)⁷

⁷ The Velleman article that Rosati references is "Is Motivation Internal to Value?"

Velleman, as characterized by Rosati, seems to say that a necessary condition of something, X, being in A's good is that A ought to care about X. Moreover, so this line of thought seems to go, a necessary condition of A's being obligated to care about X is that A is capable of caring about X. I will assume for the sake of argument that Rosati has characterized Velleman's position correctly. Rosati thinks that Velleman's position supports her internalism requirement. Let me explain.

Velleman, at least as Rosati characterizes him, is committed to claiming two theses: (1) If X is good for A, then *prima facie* A ought to care about X, and (2) If *prima facie* A ought to care about X, then A is capable of caring about X. (1) and (2), when taken together, entail (3): If X is good for A, then A is capable of caring about X (given suitable *prima facie* conditions). Now, my ideal advisor theory can accommodate (3) because (3) is merely a weak internalism requirement. Rosati offers an argument for her strong internalism requirement. I will turn to that now.

Rosati's argument for her strong internalism requirement is an argument by analogy to the "'ought' implies 'can'" principle. Rosati says:

When, in the moral case, we ask the question, "Can a person choose to die rather than to hand over her money to a thief?" we do not mean to ask, "Can she if we hypnotize her to make that choice?" or "Can she if we alter her brain?" We mean something more like this: "Can she make that choice on her own, as she is, and with all that she can muster from that standpoint?" Now, to return to the nonmoral case, there may well be things that a person would care about if only she were under conditions other than her current ones . . . The bare fact that A would care about X under C [such as A on serious hallucinogenic drugs] surely does not show that A can care about X in the relevant sense. The ought of "'ought' implies 'can'" is not addressed to the person she is under C. Insofar as judgments about our good present us with *prima facie* oughts, they are addressed to us in our actual positions. More precisely speaking, they are addressed to us as occupants of normally accessible improved conditions. (Ibid, pg. 321)

Rosati is here offering a fairly clear argument in favor of her strong internalism requirement. Roughly she sees there as being an "'ought' implies 'can'" principle and

she also sees there as being a similar “‘good’ implies ‘can’” principle. Just as the “can” in the “‘ought’ implies ‘can’” principle is understood, roughly, to be “can make that choice on her own, as she is, and with all that she can muster from that standpoint,” so too, Rosati’s argument goes, the “can” in the “‘good’ implies ‘can’” principle implies, roughly, “can care about it on her own, as she is, and with all that she can muster from that standpoint.” I think that Rosati’s argument from analogy has serious problems. I will explore a few of the problems I see for her argument by analogy.

Firstly, the “‘ought’ implies ‘can’” principle, as Rosati characterizes it, is false. Let me illustrate with an example. I, and presumably the reader, take it as given that molesting children is wrong. Yet some people cannot refrain from molesting children, on their own, as they are, with all they can muster. In such cases, child molesters are not justified in their behavior.⁸ Rosati’s construal of the “‘ought’ implies ‘can’” principle is controversial at best.

Secondly, even if Rosati is correct in her construal of the “‘ought’ implies ‘can’” principle, there are nevertheless serious difficulties with the analogy. To repeat a controversial line from the block quotation above, Rosati says: “Insofar as judgments about our good present us with *prima facie* oughts, they are addressed to us in our actual positions” (Ibid, pg. 321). Now, Rosati here makes reference to the motivational force of our *judgments* about our own good rather than the motivational force of what *really is* our own good. Exactly what Rosati has in mind here is not clear. Perhaps she is alluding to

⁸ Note that child molesters are not even excused in such cases. One is *justified* in acting a certain way if one is right in one’s action. One is *excused* in acting a certain way if one acts in a morally wrong way, yet one is, for whatever reason, not responsible for acting as he did. Rosati needs the “‘ought’ implies ‘can’” principle to provide *justification*, in moral cases, in order for her analogy to work. Yet I think she does not even get *excuses* in moral cases.

her previous argument for existence internalism from judgment internalism. As I say above, such an argument does not work. Judgment internalism, for all I know, could be true, yet it does not support existence internalism. So, to be charitable, let us suppose that Rosati, in the quotation, is writing of our actual good. There is still a problem—why suppose that there our individual good presents us with *prima facie* obligations in the first place. This is something that requires argument. Rosati seems to accept what I mention above as Velleman’s (1)—if X is good for A, then *prima facie* A ought to care about X. Yet this is something that requires argument. Simply to assume it will not do in this context. When we look at the details of Rosati’s argument by analogy, it seems that there are serious problems.

Rosati’s fifth and final argument is based on autonomy. Here is what Rosati has to say about autonomy and well-being:

Internalism is thus supported not merely by the negative concern to prevent alienation, but by a positive concern about autonomy. The negative side of insuring that a person’s good is made for or suits her is insuring that it is not something alien to her. The positive side of insuring that her good suits her is insuring that it is a reflection of her autonomous nature. We might, then, attempt to defend internalism about the good directly by appeal to autonomy. (Ibid, pg. 322)

Rosati’s fifth argument is positive in the sense that it does not argue about what does not “suit” the agent, but is instead about what does—namely, actions that reflect her autonomous action.

Rosati’s argument from autonomy goes like this:

Something cannot be a part of a person’s good if it cannot enter into her rational self-governance. And it can enter into her self-governance only if she is capable of caring about it. If she is not capable of caring about it, she cannot of her own accord rationally pursue it, promote it, or simply cherish it. (Ibid, pg. 323-324)

I take it that she means that A’s being capable of caring for something is a necessary condition of that thing’s entering into A’s self-governance which, in turn, is a necessary

condition of that thing's being good for A. Again, a lot depends on what Rosati means by saying someone is *capable* of caring for something. If Rosati just means that there is a possible world in which A cares about the thing, then the full information view can meet the internalism requirement quite easily. But presumably she does not have such a weak internalism requirement in mind.

Rosati thinks that there is a general connection between autonomy and well-being. Specifically, she mentions that one's good cannot be alien to her and what is good for her must suit her autonomous nature (Ibid, pg. 322). I agree that there might be such a connection and I think that my ideal advisor account can explain the connection. Let me explain.

There is a recent and significant body of literature in which it is argued that desires can be either autonomous or not autonomous. Whether a desire is autonomous depends on how it is formed. Inspired by Amartya Sen's writings about desire formation, many people have come to reject the thought that actual desires are invariably truly autonomous.⁹ To borrow and modify an example, we can imagine that someone is "brainwashed," through cultural inculcation, into desiring a brand new and expensive bicycle. Is this what the agent autonomously wants? It seems not. Many writers on autonomous desires are willing to grant that non-autonomous desires at least sometimes will not "extinguish" with more information, such as, for example, information about its origin. Some non-autonomous desires persist even when people find out about their origins and disapprove of them.

⁹ For a good example, see Feinberg's "Autonomy." in *The Inner Citadel: Essays on Individual Autonomy*.

Perhaps we can incorporate the idea of autonomous desires into my ideal advisor account of well-being. Admittedly, this is speculative, but we might be able to unite the “informed advisor” literature, which has its roots back in the writings of the early utilitarians, if not before, and continues on through Railton, with the literature on “autonomous desires,” which follows from Sen and the work of others. A central motivation for each of these strands of literature, I say, is that one’s personality is central to her well-being. If an agent were to have information about how a desire was formed, perhaps he would not desire his less informed self to satisfy the desire or even to have it. Perhaps an agent’s desires of this kind are his non-autonomous desires. In the previous chapter, I explored how false beliefs can distort one’s personality and result in desires that are not true to one’s personality. Such desires, I argued should not be counted as relevant to one’s well-being. Perhaps, in a somewhat similar way, cultural inculcation might distort one’s personality and result in desires that are not true to one’s personality. Strictly speaking, in the example of the bicycle above, there is not a false belief in play; rather, the distorting element is a culturally inculcated value. If Sen and others are right in thinking that desires can be non-autonomous, I think that appealing to personality is a way of explaining what they are up to. Now, a lot of what I have just said has been speculative, but I do think I have presented an intriguing possibility for an ideal advisor view. Additionally, I think the idea of the personality can unify and explain a great deal about the relationship between autonomy and well-being. I think I have sketched out a general way in which one might think that autonomy and my conception of well-being are related. In this way, I hope to have alleviated any concerns that my ideal advisor

account ends up identifying what is good for someone as alien or not true to her autonomous nature.

Let me explain why I do not pursue the details of my proposal further. I have a slight suspicion that the examples of socially inculcated values, as found in the works of Sen and others, are not as problematic for an ideal advisor theory such as mine as the examples may initially seem. Examples such as a desire for a new bicycle or the desires of a contented slave always, or at least nearly always, are morally colored in a way that might bias our intuitions on the matter. A better test case might be of a desire that is for something that is morally good, which is intrinsic and self-regarding in the way I describe in the previous chapter, such as the intrinsic desire by *A* that *A* feed the starving. Imagine that this desire by *A* is socially inculcated. Is the intuition, nevertheless, that *A*'s desire is irrelevant to *A*'s well-being? At least my intuitions become less clear in the case just discussed. I am not sure that Sen and others have identified a true class of desires that are irrelevant to one's well-being. Given that I think it is unclear whether socially inculcated desires ought to be excluded from being relevant to one's well-being, I leave the proposal for their exclusion only sketched out. My thesis that one's personality is a central determiner of her well-being can be modified to fit either side of this issue.

Let me sum up what I hope to have shown thus far in this chapter. Rosati and Loeb think that there is an internalism requirement. Understood in a very rough way, the internalism requirement links one's good with one's motivation. A weak version of the requirement is that in order for something to be good for someone, it must be possible, in a broad sense, for the person to care for it. My and Railton's ideal advisor views meet the weak internalism requirement. A strong version of an internalism requirement is Rosati's

two-tiered view, according to which, roughly, X is good for A only if (1) A_{ic} cares about X for A and (2) A_o cares about A_{ic} 's caring about X for A . None of Rosati's five arguments supports her strong internalism requirement. What I will do now is offer some arguments against Rosati's strong internalism requirement.

Rosati's specification of the requirement does not itself remain true to existence internalism. Recall that existence internalism, in the words of Darwall, is the theory, roughly, that there is a necessary connection between the truth of a normative proposition and motivation (Darwall, pg. 308). I will now adjust Darwall's characterization of existence internalism to fit into a discussion of well-being. Roughly, existence internalism about well-being is the theory that there is a necessary connection between what is prudentially good for someone and his motivation. I certainly do not wish to commit myself to existence internalism. It is a very controversial thesis. I will argue, however, that even if existence internalism were right, it would not be entailed by Rosati's strong internalism requirement.

Rosati's characterization of the internalism requirement—the strong version—does not make what is good for someone depend on that person's motivation to pursue it in cases where she does not have it.¹⁰ According to Rosati's construal of the requirement, while A_o has to care about what A_{ic} advises, A does not. So something can be good for A without A having any motivation to pursue it. That aspect of Rosati's position might seem surprising. But let us set aside this worry.

¹⁰ In this passage I will write of desires as though they are always for things that one does not have. This is false, of course. People may desire things that they have; perhaps such cases might better be called cases of cherishing rather than desiring, but that this beside the point. I will opt for simplicity of expression.

There is yet another “gap” between what is good for someone and his motivation to pursue it. Rosati does not require that A_o care about what A_{ic} advises: instead, she requires that A_o care about A_{ic} ’s advice. Those are two very different things. If we think of the advisor’s advice as being a kind of desire about A , then A_{ic} ’s advice will consist in a set of desires. Now, A_o does not have to have the same desires as A_{ic} . Instead, A_o merely has to care about A_{ic} ’s advice. Let me put this more carefully. Let us say that A_{ic} ’s advice for A consists in a set of desires for A to have a variety of things, x . A_o does not have to desire x , he can merely care about the advice of A_{ic} , which means A_o might have no motivation to pursue x whatsoever. Here is an illustration by example. A_{ic} might advise A to improve his health. A_o might approve of the fact that A_{ic} has this advice for A . Nevertheless, A_o might not desire to improve his health himself. I could provide many similar examples. Rosati’s internalism requirement, if I am right, does not always require that an agent have some motive for his own good: it merely requires that one have a positive attitude toward an ideal counterpart’s advice.

So I do not think that one of the core ideas Rosati says motivates her strong internalism requirement, existence internalism, supports her two-tiered version of the requirement. It is hard to see what else supports her strong internalism requirement. Railton’s and Velleman’s sparse comments on this issue are far too inchoate to support Rosati’s strong internalism requirement. Attempting to establish the strong internalism requirement is a very ambitious project. If strong internalism were true, then virtually all popular accounts of well-being would be wrong. Thus, it is not surprising that the strong internalism requirement is incredibly hard to establish. It is not clear whether the weaker

internalism requirement is true. However, even if it is true, my ideal advisor theory meets the requirement.

Perhaps the reader will be surprised to find me in agreement with Rosati on one “big picture” issue regarding desire-satisfaction accounts of well-being. Let us say that an “actual” desire-satisfaction view is that one’s well-being varies with the extent to which one’s actual, intrinsic, self-regarding, non-moral and non-aesthetic, desires are satisfied. Rosati is right to be concerned about the central motivations for the move from the actual desire-satisfaction view to ideal advisor accounts. Rosati seeks to locate the central motivation. So do I. She finds the central motivation in her strong internalism requirement. I, following Railton in a way, find a central motivation in the requirement that relevant desires to one’s well-being fit one’s personality. So, Rosati and I agree that a central motivation must be found. We just disagree about the nature of the central motivation.

Ideal Advisor views have been criticized because they do not meet an internalism requirement and I take myself to have discharged that criticism. There are, however, other criticisms of Ideal Advisor views. I will turn to them now.

4.3 Full Information and the Conditional Fallacy

As I explained in the previous chapter, the ideal advisor view I endorse requires the advisor to have full as opposed to partial information. The position is counterfactual in that it does not require a fully informed actual person to want a kind of life for A. All that is required is that A’s being fully informed is in principle possible. Counterfactual analyses have a danger attached to them. Any analysis fails when some unintended cases are captured by the analysis or some of the intended cases fail to be captured. Robert

Shope, in his paper “The Conditional Fallacy in Contemporary Philosophy,” explains a special way in which counterfactual analyses can fail.

The language of possible worlds helps make counterfactuals clear. Shope did not use the language of possible worlds when he discussed the conditional fallacy, but the fallacy, when stated in the language of possible worlds, is much easier to grasp than it would otherwise be. So, I will characterize the conditional fallacy in the language of possible worlds.

Imagine that person *A* is in the actual world. Imagine that *A*⁺ is *A* in the nearest possible world in which *A* is fully informed. By the “nearest” world I mean the world in which the fewest possible facts change and yet *A* is fully informed. There is no guarantee that the two possible worlds differ in just the one fact about the information that *A* has. Once we look to the nearest possible world in which *A* is fully informed, there is no guarantee that *other* things are not different as well. Let us call these other differences “consequent changes” because they are not intended differences from the actual world. Rather the nearest possible world in which the specified, intended, conditions are met also contains these consequent changes.

Some consequent changes will not cause problems for an analysis and some will. People speak of Cambridge changes.¹¹ A Cambridge change occurs for some object when there is a change in the statements that are true of that object.¹² For example, assume that George Bush cut his cheek while shaving today. There is a substantial change in Bush in that he now has a cut on his cheek. There is also a Cambridge Change

¹¹ The discussion in this paragraph is drawn from Parfit in his “What Makes Someone’s Life Go Best” in *Reasons and Persons*.

¹² This is the definition of a Cambridge Change as found in Parfit, pg. 494.

in Aristotle. Aristotle lived in a world in which, later, Bush cut his cheek. If we adopt the Cambridge change example to the ideal advisor examples, there will be lots of Cambridge changes from the real to the nearest possible world in which *A* is fully informed. For starters, everyone will bear a relation to *A*+ instead of *A*. Cambridge changes will not cause problems for counterfactual analyses but at least some other changes might.

One example of a consequent change that could cause problems for the full information account of well-being is that *A*+, now with full information, may dislike *A*. *A*+ could find *A* to be a despicable sort of fellow. *A*+ could then advise *A* to jump off a cliff.

Peter Railton has offered a way in which the above-described problem can perhaps be met. Using Railton's language in construing ideal advisor theories, we place the further constraint on what *A*+ can want for *A* to want.¹³ Railton says the following: "The wants in question, then, are wants regarding what he [*A*+] would seek were he to assume the place of his actual, incompletely informed and imperfectly rational self [*A*], taking into account the changes that self is capable of, the costs of those changes, and so on" (Railton, "Facts and Values", pg. 16).¹⁴ Railton thinks that we can make it a requirement that *A*+ choose as if he were to take the place of *A* in the immediate future.

¹³ In this passage, I will stick with the desire to desire language as a way of phrasing the advice of an ideal advisor that is relevant to well-being. However, please bear in mind that I think we should drop the desire to desire language altogether. Please look back to Chapter 3 for the complications concerning this issue. For the present, I will use the desire to desire language for simplicity.

¹⁴ Railton, in this quotation uses the "desire to desire" terminology referred to above. As I note above, I think that the "streamlined" account of ideal advisor theories is superior in capturing the central ideal of ideal advisor theories. In this context however, I will stick with the "desire to desire" terminology for simplicity.

Although Railton's proposal resolves some of the counterintuitive results, it does not seem to solve a central worry. The original problem was that it seems possible for the fully informed self to cease to care about the ignorant self. The original problem exposed the unjustified assumption that we care about ourselves under different possible scenarios (what we take to be different possible worlds). Railton's proposal is that we treat the fully informed self as thinking as if he were to adopt the life of the relatively ignorant self in the immediate future, but the basic problem still remains in that A+ might still think something such as "if I were to be as A is, I would want myself to jump off a cliff."

A thought of Darwall's will help here. As is discussed in my Chapter 1, Darwall thinks that a person's well-being is best understood as what we would rationally want for her for herself, were we to care for her. As I have said, I think that Darwall is writing here at the level of concepts, not conceptions. Nevertheless, his idea can be brought to the level of conceptions. We can solve the problem under discussion by including the requirement that the fully informed self care for the actual, less informed, self. That solves the original problem because it forbids the cases where we do not care for our future life paths (which Railton's proposal does not rule out).

Darwall, of course, thinks that anyone, if she were to care, could give advice to someone and that the content of that advice would be constitutive of her advisee's well-being. My position is different in that I take the standpoint of the agent himself to be privileged over those others who might happen to care for him. The reason is that others might care for someone deeply, but not really know the person all that well on a deep level or even have a large commonality of dispositions and past experiences. In this case, it seems clear, their advice would not fix the person's well-being. As the reader might

expect, my reason for preferring my account over Darwall's has to do with considerations stemming from the relation between one's personality and what is good for one. Advice from one with the same personality as his advisee, I say, is authoritative in determining what makes the advisee's life go best.

David Sobel, in his "Full Information Accounts of Well-Being," offers another example of a consequent change that seems to threaten the full information view. Sobel worries that the fully informed self might be driven mad after undergoing a process through which he gains full information.¹⁵ Ignoring for the moment details about the process (I will have more to say about that later), Sobel's concern looks a lot like the worry about whether the fully informed counterpart would care for his less informed counterpart. Sobel does not purport to show that someone who went through the process of being fully informed would go mad. Instead he only has a concern that someone might become mad after undergoing the process. This would be an undesirable consequent change. This worry can be countered by adding the requirement that the person must be sane after the process of gaining full information.¹⁶

I think that so long as the conditional fallacy has the form of "for all we know, x could be a consequent change, given your analysans, and if so, your analysis will yield counterintuitive results," then a solution can be provided. I am optimistic here because although sometimes critics of ideal advisor accounts write as if there would be only one possible world in which one would have full information, and they then prognosticate

¹⁵ Think back to Brandt's description of what goes on during cognitive psychotherapy: "persons repeatedly represented to themselves, in an ideally vivid way, and at an appropriate time, the available information . . . the person gets the information at the focus of attention, with maximal vividness and detail, and with no hesitation or doubt about its truth" (Brandt, pg. 111-112). It does seem possible that some people would go mad after undergoing such a process.

¹⁶ Again, how the process of being fully informed would go will be discussed later in this chapter.

doomsday-like scenarios, in reality (in some sense perhaps), there are many possible worlds in which one is fully informed, i.e. there are many *A+* candidates. Of these many possible worlds in which one has full information, some will involve changes in one that will be problematic, but some will involve few significant changes in one or changes that are quite unproblematic. The project, as illustrated above, would be to single out the possible world in which one is fully informed that works best in the theory. The criterion by which one evaluates which of these possible worlds is best for our purposes has to do with the basic motivation for desire-satisfaction views. If I am right in thinking that one's personality is central to her well-being, then the concept of a personality should be a key factor—the desire-set of the *A+* who reflects the full range of *A*'s personality and does not distort *A*'s personality is the desire-set relevant to *A*'s well-being. As I explained in the previous chapter of this dissertation, if an advisor has *incomplete* information, then his advice might *not reflect the full range* of one's personality and if an advisor has *false* beliefs, then his advice might *distort* one's personality. My present point is that additional restrictions may need to be placed on *A+*'s advice.

Problems that can be solved by appeal to the concept of a personality include worries about the increased mental capacity and mental processing speed that would have to be the case in order for someone to have full information. To be sure, it is a little odd even thinking about people with what can only be described as super-minds. However, anyone's personality, I say, is compatible with increased cognitive capacity and the like. Unless someone can show that some counterintuitive consequent change *must* occur in a world in which one is fully informed, the method of singling out the best possible world in which one is fully informed seems to have a great deal of promise.

My ideal advisor account of well-being treats advice as relevant if it is from one's counterpart in the nearest possible world in which the counterpart is fully informed and shares the same personality. Thus far no one has come close to offering an argument that counterintuitive consequent changes must result in the nearest possible world in which one is fully informed. There could be many possible worlds, perhaps infinitely many possible worlds, and my account of well-being involves picking out just one.

I will now turn to several criticisms of ideal advisor views based on the very idea of one's being fully informed.

4.4 Knowledge That and Knowledge What Something is Like

Take knowledge that something is the case to be "propositional knowledge." It is factual and everyone who is fully informed would have the same knowledge of this kind. Examples of propositional knowledge would be that the Earth is round and that the sky is blue. Knowledge of "what something is like" is different. I have in mind here knowledge that is not propositional; the object of the knowledge is one's experience of the thing in the world, or perhaps the thing in the world. Examples of non-propositional knowledge would include knowledge of what pineapple tastes like and knowledge of what teaching a philosophy class is like. There at least seems to be *prima facie* reason for thinking that the two kinds of knowledge are distinct and that one kind is not reducible to another. This is a heavily debated issue, of course.¹⁷ If there were only propositional knowledge, then full information accounts of well-being would have a relatively straightforward explanation for how someone could be fully informed. The issue is far

¹⁷ One recent excellent collection on the issue is *There is Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument*.

more complicated if there is non-propositional knowledge. For the sake of argument, I will grant that there is a substantive, non-reductive, kind of non-propositional knowledge.

Rosati in a different paper than the one discussed above called “Persons, Perspectives and Full Information Accounts of the Good” and David Sobel in his “Full Information Accounts of Well-Being” offer different arguments that draw in question the plausibility of including knowledge of what things are like in the set of full information. Indeed, they claim to challenge the very intelligibility of an agent’s being fully informed.¹⁸ First I will explore the plausibility of including knowledge of what things are like in the information set. Ultimately, I think that at least some knowledge of what things are like must be left out. Later, I will examine whether full propositional information is adequate as an information set for the ideal advisor. While Rosati and Sobel are skeptical, I argue that full propositional information is far richer and more textured than they envision and may very well be sufficient to play the requisite role in the deliberation of the ideal advisor.

Rosati has two primary lines of argument against ideal advisor theories in her paper “Persons, Perspectives and Full Information Accounts of the Good.” The first is the argument that the fully informed self will not necessarily be someone whose judgment

¹⁸ God is said to be all-knowing. One really fascinating thought is whether Sobel’s and Rosati’s arguments purport to show that God could not be fully informed in the required way. The agent with full-information will be God-like *in a way* on the full-information ideal advisor accounts. So it seems like the two cases of being fully informed might have to stand or fall together. Sobel’s and Rosati’s arguments would prove something quite extraordinary, incredibly significant, if they showed that one of the three supposed properties of the traditional, Western God is not intelligible. Another interesting issue is whether God could play the role of the ideal advisor. Carson, in his book *Value and the Good Life* proposes just this. Carson seems to think that God exists and runs his theory from there (though in fairness to Carson, he also runs a parallel argument in case God does not exist). However, no one has to presume that God exists. For, there is a God of the relevant sort in some possible world, even if not in this world. Why is his advice not authoritative for each of our welfares (the advice he would give if he were to exist, for those who are not possible worlds realists)? There are many interesting issues that arise in theology. But there are also serious issues that are central to my dissertation.

you recognize as authoritative. This line draws heavily from her earlier paper on internalism and in fact presupposes the internalism requirement. Because I take myself to have rejected her version of the internalism requirement as a requirement on well-being or the individual good, I will not deal with her first line of argument.

The second primary line of argument has the conclusion that no person can be fully informed. Rosati summarizes her thinking on this in the following quotation:

Part of being a particular person with particular traits is occupying a point of view—one that involves a certain way of seeing, feeling, and evaluating and which gives access to certain information while making other information inaccessible. If a person is to be fully informed, however, she must be able to enter into all her possible points of view. She must be capable of now appreciating all her lives as the persons she would be if living them—side by side, so to speak. The problem concerns how she can occupy a point of view that gives her equal access to viewpoints that may be in direct conflict, each excluding information accessible from the other. (Rosati, 1995, pg. 317)

Roughly put, Rosati's argument seems include two thoughts: (1) that people have to occupy a single perspective in order to make judgments at all, and (2) that being fully informed, in the relevant way, requires experiencing things from different perspectives, which is impossible.

Rosati gives an example of experiencing things from the distinct viewpoints of an obtuse person and a sympathetic person. "But when it comes time to compare these lives, a person must have features that enable her to appreciate both what it is like to be her obtuse self and what it is like to be her sympathetic self" (Ibid, pg. 318). No such feature can be found, Rosati argues. The fully informed self cannot herself be both obtuse and sympathetic, for they are mutually exclusive (Ibid, pg. 319). The agent who is fully informed will be either obtuse, sympathetic or some other way, but for each way she turns out, there are countless other ways that are excluded.

Rosati is not quite right in thinking that there is no possibility that the informed person could compare the two ways of living. The fully informed self would not have to *be* both obtuse and sympathetic. Perhaps full *propositional* knowledge of how the two different ways of living compare would be sufficient for comparison. Or perhaps one can have knowledge of what it is like to be obtuse and knowledge of what it is like to be sympathetic without actually being obtuse and sympathetic.

There is a further objection to the way Rosati construes the problem for the full information account. I think that the problem Rosati puts forth is not one that strikes at the core idea behind the full information theories of well-being. Let me explain.

I think that Rosati's example is poorly constructed. I will go back to the way in which she describes the example and problem for the ideal advisor theory. I do not think that knowing what it is like *to be a certain kind of person* is a central test case. It looks like Rosati is demanding that the ideal advisor account provide an explanation of the "birds-eye" point of view that can incorporate knowledge of all of the different personalities a person could have. I do not think that the ideal advisor account owes such an explanation, for the ideal advisor account does not even allow that the personality of *A+* be different from the personality of *A*. Remember that I argued that the best versions of ideal advisor views treat the personalities of *A* and *A+* as constant. Now, some possible desire sets regarding *A*'s life will be "truer" to *A*'s personality than others. But a proponent of the ideal advisor theory I favor does not think that there is a perspective outside one's personality from which to judge which sort of person to be (e.g. obtuse versus sympathetic). So I think that Rosati is demanding an explanation that a proponent of ideal advisor views both does not owe and should not try to give. Rosati would have

done better to put the issue in terms of knowing what different experiences are like from a *specific* perspective.

Now, I think Rosati is clearly on to something with her arguments. However, I think that the real issue though is knowledge of what two mutually incompatible experiences are like, from a given perspective. Sobel captures the problem nicely and I will soon move to his discussion soon, but I must first finish the discussion of Rosati's argument and its implications.

I agree with Rosati that, given the nature of what it is to occupy a single perspective, there is some conceivable information that cannot be included in the information set with which the ideal advisor deliberates—namely, information that is only accessible to those with a different perspective. However, I disagree with her about the implication of that fact. She thinks that this undermines full information ideal advisor views of the individual good. I think it does not. Part of what has restricted full information accounts all along has been a restriction on what information it is possible to have. I agree with Rosati that a singular perspective is required of an ideal advisor and that it is impossible for the advisor to occupy another perspective. I accept that there is an intelligible kind of information that cannot be included in the set to which the ideal advisor has access; for any perspective, there is some information that a person with that perspective cannot have. If that means that my account is not genuinely a full information account, then I am willing to accept that. So my position could be called an “as much information as a personality (and a perspective) will allow” view. I believe, however, that it is only fair to think that full information has been constrained all the while by what is possible given the nature of perspectives.

Sobel, as I stated above, captures the challenge to ideal advisor views in a clearer fashion than does Rosati. Sobel's challenge is similar to Rosati's in that it involves knowledge of what things are like. Sobel explores two different ways of thinking about the full information view: the report model and the experiential model. On the report model, full information is obtained when one learns of and comprehends reports about the various possible future scenarios. On the report model, the agent lacks firsthand experience with what the world is like (Sobel, pg. 796). In my terminology, fully informed agents on the report model have only propositional knowledge that things are the case. Sobel argues that the report model is inadequate, and I will come back to visit this question in a little while.

The experiential model of full information is the model that gives the agent firsthand experiences of what things are like. This model breaks down further into two distinct types: the serial model and the amnesia version. On the serial version of full information, an agent experiences a specific set of events and then later goes on to another set, and so on. A simplistic example is the process of sampling various flavors of ice cream in order to decide which one to buy. Not all kinds of experience will fit the simplistic example, so the argument goes. Some experiences will make it so that the agent cannot experience another set of circumstances. Sobel's example is of the various ways in which one's first kiss might go. If one experiences his first kiss one way and then moves on to an experience of the next, the second case cannot be of a first kiss. In this way, it is impossible for one informed serially to be fully informed of all of the ways of life open to one (Sobel, pg. 803-805).

The amnesia version solves the problems that the serial version faces, supposedly, by making it so that the agent experiences something and then forgets it, experiences the next thing, and then forgets it, and so on. Later, the agent comes to remember all of the various experiences that he or she had as if these experiences were for the first time. Sobel argues that this version is still inadequate because it fails to give a single unifying perspective from which to judge the various kinds of lives. On this topic Sobel writes:

For the purposes of the full information account we need a perspective whose preferences between the lives in question accurately determine their value to the agent. But now a problem arises. Our actual evaluative perspective changes over time. We can therefore expect that we would respond differently to factual and experiential knowledge at different times in the future. Thus we do not have a single informed perspective to deal with, but several. And each will offer occasionally conflicting assessments of where the agent's well-being lies. (Ibid, pg. 805)

There seem to be three different worries in this passage: the first is about the plausibility of the amnesia version of full information, the second is about whether there is a single evaluative perspective for any agent at a specific time and the third is whether there is a single evaluative perspective over time.¹⁹

I think that the third problem Sobel mentions is off target if thought of as a problem unique to full information accounts of well-being. Indeed it applies at least to all plausible desire satisfaction theories, and to a very broad range of accounts of well-

¹⁹ A fourth worry that might be in the background of Sobel's remarks is that the *order* in which one experiences things might affect how they strike one. I am not sure whether Sobel does indeed have such a concern for the amnesia model because I am not sure whether Sobel thinks that the order of experiences could even be a problem for the amnesia model. The way Sobel thinks of the amnesia model, for the patient undergoing the experiences, it is as though each experience is without the others—for the first time. Of course, there could be a real order, but the patient, even after remembering all of the experiences, might not be able to tell the order on the amnesia model. I will not pursue this issue because I think that it leads us off the central thread of what Sobel is trying to argue. In any case, since I think that the amnesia model is ultimately doomed, I will not try to defend it too vigorously.

being.²⁰ So long as we fix the desire set to the current desire set (or the set we would currently want ourselves to want were we currently fully informed), then at any given moment, well-being can be determined in principle. Indeed, in the first chapter of this dissertation, I note that the account of well-being that I am after is an account of one's well-being at a time. I will not re-open the issue at this point, but I refer the reader my discussion of Velleman's thoughts on well-being.²¹ The problem of temporal perspectives is not unique to full-information accounts, nor even unique to desire-satisfaction accounts, but applies to all theories that allow something to raise well-being at one time which would lower it at another.

Let us now look at the second problem Sobel seems to have in mind, where we keep the time constant and examine whether there is a single perspective from which to judge different possible life paths. Let us treat the amnesia version of how one might acquire full information as unproblematic, at least for now (later I will have more to say about it). *I think that whether there is a single perspective from which to judge the various life paths depends on whether a single personality survives the process of being fully informed.* As I stated above when I discussed Rosati, I am willing to admit that, given the nature of a personality, for any personality, there is some information that a person with that personality cannot have. We ought not to think of the information set as including knowledge from radically different evaluative perspectives. Some information is excluded: only information available given the nature of the personality is admitted into the information set with which the ideal advisor deliberates.

²⁰ Brandt has argued this and although Sobel *does* mention Brandt, Sobel does *not* mention that Brandt thinks that this is a problem for all desire-satisfaction theories.

²¹ I will discuss this issue again in the final chapter.

Now, I will turn to the first problem. The first problem is about the plausibility of the amnesia version of being fully informed. Admittedly, the amnesia version has some unattractive features. Sobel sees two things wrong with it:

First, it would surely require much very complicated research to have anything to say about the question of the similarity of experience between the original having of an experience and a later sudden recollection of it in cases in which the evaluative perspectives held at the time of the original experience differs significantly from that held by the person who is suddenly remembering. Second, even if we can convince ourselves that some system of controlled amnesia would enable us to avoid the problems I have mentioned above, there is no way of estimating the psychological shock of experiencing such a large number of instances of amnesia and loss of amnesia to our idealized selves. The full information theorist cannot simply stipulate that the idealized agent remain sane through this process, and I would have concerns about this issue. (Ibid, pg. 807)

Sobel has two substantial worries about the amnesia version of full information.

But I wonder if the amnesia and the serial versions are the only options available.

Moreover, Sobel, I take it, thinks that a proponent of an ideal advisor view needs to provide an account of the process by which one comes to have full information. I do not think that is so.

The “possible worlds” analysis we have been working with has depended at least on what is *logically* possible. Everyone agrees that it is not physically possible given scientific laws, for a person to become fully informed. This much has always been taken for granted by the proponents and opponents of ideal advisor views. The sense in which it is possible for one to be fully informed, is thus logical, but not physical. Miracles are logically, but not physically possible. There will be some possible worlds in which A+ becomes fully informed by miracle and some where he becomes informed by a process such as the amnesia version as described by Sobel above. The world where A+ is informed by miracle would likely be one in which there are fewer problematic

consequent changes than the world in which he undergoes the amnesia process. So we should think of A+ as being informed by miracle.

The reader may find my resort to the idea of A+'s being informed by miracle to be *ad hoc*. I, on the contrary, find it hard to see how one could be fully informed in any other way than by miracle, on my account. My account of well-being involves identifying the nearest possible world in which one's counterpart is fully informed. Any agent who undergoes any of the processes Sobel mentions will *not* be in the nearest possible world in which the agent is fully informed to the original world. One's fully informed counterpart can differ from her actual self only in being fully informed, and one's counterpart cannot have gone through any of the processes Sobel proposes, for if one's counterpart were to go through such processes, she would *not* be in the nearest possible world in which she is fully informed. Provided there is a world in which one is fully informed by miracle, it is the nearest world in which one is fully informed.

Also, I have a deep concern about how Sobel characterizes unique experiences, such as first kisses. Presumably, the phenomenological experience of what something is like does not have a "firstness" or "secondness" built into it. Instead, the phenomenology of a first kiss and a second kiss, where they are different only in order and other prerequisite changes, is the same. Sobel must have in mind that it is a *cognitive component*, such as the belief that one kiss is first and one second, that differs. I think that there is good evidence that Sobel must think this. For example, in the example Sobel uses to motivate or explain the serial version, one is not in fact kissing different people for the first time—or the same person in different circumstances; instead, one is kissing different people with the *belief* that it is one's first time. Sobel, and I think Rosati too,

must think that the background cognitive components are crucial to experience or at least to the reaction to such experiences. I am not convinced that the background cognitive components are crucial. To be sure, there has been no argument on either side.

However, I feel that the burden of proof lies on Sobel and Rosati to explain how the cognitive components work in affecting our experience. How it is that the background cognitive component works, at this point, is mysterious. Moreover, perhaps the differing backgrounds for any experience are something that mere propositional knowledge could account for. I do not intend to decisively refute Sobel or Rosati on this point. Rather, I merely feel that there is a debt of explanation on their part as to what the background cognitive component is and how it functions in experiences, specifically, experiences about what things are like.

Let me briefly sum up my defense of full information theories against the concerns of Rosati and Sobel. Rosati thinks that it is impossible to be fully informed given the nature of a perspective. I agree that one's personality might not allow for some knowledge, but I do not think that this is at all problematic because, all along, my and Railton's ideal advisor theory have been motivated by the idea that one's personality is central to her well-being. Sobel's arguments cause more serious problems for an ideal advisor view than Rosati's, I think. Sobel has three concerns, the last of which is the most serious. I have responded along two lines. First, a proponent of a full information view does not owe an account of the process by which one becomes fully informed. Given that full information views have always been about what is logically possible, rather than psychologically possible, a proponent of a full information view can stipulate that one is fully informed by miracle. Second, I think that Sobel owes an explanation of

how beliefs shape experiences in the way he seems to think they do. Sobel's presuppositions about the role of belief are controversial.

Knowledge of what things are like is part of the information a fully informed individual would have, for my ideal advisor account. Now, if some information is beyond what any one individual could have, given his or her perspective, then her ideal advisor will not be fully informed in an extremely broad sense. However, I should note that this informational limitation is not as problematic as Rosati and Sobel make it out to be. Since the ideal advisor account of well-being I have developed requires that one's ideal advisor have the same perspective as her advisee, and her advice be for her advisee given her advisee's personality, then the missing information seems irrelevant to the content of the advisor's advice. It simply does not matter what something is like from someone else's perspective, I say, given that the advisor's advice is offered from her own perspective. It seems to me, given the arguments I have offered, that Rosati's and Sobel's arguments about the very idea of being fully informed do not present serious difficulties for my ideal advisor account.

Suppose I am wrong about my arguments involving knowledge of what something is like. I nevertheless think that my full information account of well-being does quite well with merely propositional knowledge. In the part that follows, I explore the prospects of an ideal advisor account of well-being on the assumption that ideal advisors can only be informed with propositional knowledge.

4.5 Propositional and Non-Propositional Knowledge Revisited

Recall that Sobel discusses the report model in which only "knowledge that," or propositional knowledge, is permitted into the full information. He argues that the report

model has serious deficiencies, but I do not think he is right. View the work in 4.5 as a way *covering my bases* if my arguments in 4.4 do not work.

Sobel first looks at what he takes the report model to lack. Here is what he has to say:

We often need firsthand information if we are to be ideally situated, and if we are to be assured that we are ideally situated, to measure the value of a life to ourselves. . . . Some experiences are revelatory in the sense that they alter our responses to facts and descriptions. Revelatory experiences enliven our appreciation of facts and descriptions such that although we were vividly aware of the facts and descriptions of the case, we had previously been dead to the import that we now find in them. (Sobel, pg. 797)

Later, Sobel makes a key move in his arguments against the report model when he begins to speak of the experiential model as incorporating the *evaluative perspective* and the report model as failing to do so. On this Sobel writes:

Perhaps there are lives such that almost no matter which of our possible evaluative perspectives we have, we still see that those lives are wonderful or horrible for us to live [the example Sobel gives is of life in a concentration camp]. But there are surely many lives that we could lead in which our evaluative perspective in that life plays a large role in how we would experience that life from the inside. In many of these cases the way the life feels from the inside will be an important factor in determining the relative value of that life. I therefore see little reason to be optimistic about the attempt to capture the worth of these lives when we are provided only with the facts and experiences accessible to a perspective outside the evaluative perspectives which would be ours in the life in question” (Sobel, pg. 799).

Sobel does not describe the evaluative perspective in more detail. I will look at a few different interpretations and I will argue that the report model does not fail to capture the evaluative perspective in an important sense of “evaluative.”

At least in any important sense of “evaluative,” the report model will capture what is important. To go back to the earlier example about the first kisses²²; the report model

²² Recall as well that I have worries about what is going on in such cases. I take the experience of what something is like to be just the phenomenological qualities one has and I think that there is not a uniqueness

will be able to capture whether someone *regrets* or is *overjoyed* at having had one's first kiss with such and such a person at a particular place and time. Regret, resentment, disappointment, indifference, delight and joy can be propositional attitudes that are expressed in the contents of propositions. "I am delighted that Sophie made it to rugby practice on time" expresses a proposition about one's delight. Knowing what it is like for Sophie to be on time or what it is like to be delighted that Sophie is on time is secondary to the fact that one is delighted that Sophie is on time. The report model captures the important evaluative information. In fact, I have a hard time seeing how knowledge of what something is like is evaluative in any important sense—or at least what the content of that evaluation is that cannot be captured in a proposition. Evaluative attitudes are expressible publicly. So at least in one way, an agent with a single personality and full propositional knowledge has an evaluative perspective. Perhaps Sobel has a different sort of evaluative perspective in mind.

Sobel has strong concerns about the commensurability of very different experiences.²³ For example, the ideal advisor, A+, may well know that A is happy with a first kiss and happy with a different case of a first kiss. How, Sobel wonders, can these cases be commensurable (Ibid, pg. 784-786)? The report model seems to take it for granted that the reports (which are sets of propositions) about the various possible A's who undergo the experiences will make their possible lives commensurable. Two replies are in order. First, it is not clear that the experience model, even if it could be made to

such as "firstness" built into them. Instead, the uniqueness is reflected in the content of the person's background beliefs at the time of the experience.

²³ Distinguish between interpersonal commensurability, which is between different persons, and intrapersonal commensurability, which is within one person. Much has been written on interpersonal commensurability. Sobel's project in his paper is to raise concerns about intrapersonal commensurability on full information ideal advisor views of well-being.

work would help with intrapersonal commensurability. We would have these different possible sets of experiences each of which is incommensurable. It is worth noting that at least some of the other conceptions of well-being also have this problem. Happiness should probably be thought of as consisting in many different mental states. How these are commensurable is anyone's guess. Objective list theories have a similar, if not deeper, failing. Perfectionist theories have a similar problem as well because it is not clear how the differing essential components of a life are commensurable (e.g. developing one's rational nature and developing one's physical nature). Secondly, Sobel demands a very extreme kind of commensurability, one that our ordinary everyday decisions do not meet. The vividness of detail Sobel seems to require for making decisions is very difficult to have; this, even for decisions based on knowledge of what things are like from one's past experience. Even simple decisions will lack this kind of vividness of information necessary for commensurability on Sobel's view that we encounter in everyday life. This makes me think that the kind of commensurability that Sobel is after might not be all that important. It is a kind of deep commensurability that humans rarely have in normal lives and does not affect choiceworthiness in a life. I think that the report model is less flawed than Sobel suggests.

Just by way of summary, I do not think that an ideal observer theory must rely on the report model, but I do think that the report model is less flawed than Sobel seems to think. I think that my arguments in 4.4 work to show that a lot of non-propositional knowledge—all of the non-propositional knowledge that is relevant to one's advice—will be in A+'s information set. My attempts to show that the report model is less flawed than Sobel suggests is to "cover my bases" in case my arguments from 4.4 fail.

4.6 Railton's Account of Personality

I noted in Chapter 3 that Railton has his own account of the role of the concept of personality in desire-satisfaction accounts of well-being. I should note that while I am influenced by his account of personality, I differ with him significantly about its role in desire-satisfaction accounts of well-being. Let me explain.

Just before Railton introduces his concept of personality, he discusses the process by which one becomes fully informed (Railton, "Facts and Values," pg. 57-60). He is concerned with how such a process could go. As I argue above, however, I think that we do not need to provide an account of such a process—indeed, it would be incorrect to think that the fully informed counterpart of oneself would have become fully informed by going through a process of any significant kind, other than a miraculous process, since if he were to undergo such a process, then he would not be in the "nearest" possible world to oneself. Yet Railton is led to propose an account of personality that is similar to mine in response to what he believes to be problems in explaining the procedure by which one's fully informed counterpart becomes fully informed. Here is the relevant passage:

[A personality] is a collection of properties that ground dispositions to react in various ways to exposure to certain facts. Just as there is a reduction basis for an individual's current desires—those features of his psychology, physiology, and circumstances in virtue of which he now has these desires—there is a reduction basis for his idealized hypothetical desires. When we ask how his desires would change upon the impact of further information, we appeal to this basis. We, in effect, hold this basis as nearly constant as possible when asking what someone like him would come to desire—or, more precisely, would come to want that he pursue were he to assume the place of his original self. (Railton, "Facts and Values," pg. 60)

Let me assume, for the sake of simplicity in this discussion, a dispositional account of desires and personality. That is, let me assume for the sake of argument that desires are dispositions of a certain sort and that one's personality is a set of dispositions of a

certain sort. Given this assumption, then, the notion of personality that I have in mind in my account is, roughly, a set of dispositions to have desire-dispositions. On my account, *A* and *A*⁺ share a personality, and this is one of the things that makes *A*⁺'s advice authoritative for *A*'s well-being. But Railton seems to have something different in mind.

A person's personality, on Railton's account, seems to be the reduction basis for his dispositions to have dispositions. Railton, then, has a very different understanding of what a personality consists in, it seems, than I do. Railton seems to have in mind that when *A* gets further information, an important part of an ideal advisor account of well-being is that it keeps *A*'s personality the same, which is to say that it keeps the reduction basis for crucial dispositional elements of *A*'s psychology the same. That much I can agree with, but Railton's use of the concept of personality is importantly different than mind. For on my view, what is crucial is just that the relevant dispositional elements of *A*'s psychology remain the same.

If I am right in my interpretation of Railton, his notion of the reduction basis of relevant dispositions plays a *somewhat* similar role in his account to the role that the idea of personality plays in my account. In my account what is important is to keep the personality as constant as possible between *A* in the relevant two possible worlds.

Railton and I differ in our accounts of personality and its role in an ideal advisor account of well-being. In Railton's usage, a "personality" seems to be the reduction basis for what I refer to as a "personality." I do not here have an argument that he misuses the term "personality." Nor do I have an argument that there is only one concept of personality and that mine is right. The important point, I think, is that he and I are talking about different things. I think that the term "personality" captures nicely what I have in

mind and I will continue to use it. I can agree, of course, that there may be a reduction basis for what I refer to as a person's personality, but it is the personality that plays a role in my account, not its reduction basis.

4.7 Conclusion

If I am right about what I say in 4.4 of this chapter, $A+$ will have full propositional knowledge and as much knowledge of what things are like as is possible, given her personality. Although Sobel and Rosati worry that the lack of absolutely total information is problematic for ideal advisor theories, I find that it is consistent with what I take to support ideal advisor theories all along. I think that one's personality is central to one's well-being. It is not surprising that the very having of a personality limits the information one can have, given that each personality occupies a single perspective.

At this point in the dissertation, we have a more complicated possible worlds analysis of my preferred ideal advisor theory. Let me take a step back and go over the full theory with all of the modifications introduced since the previous chapter. There will be lots of possible worlds in which A is fully informed and is a candidate to be considered $A+$. We must now pick out a subclass of the set of all such candidates to be $A+$. The ideal advisor must care about the relatively less informed A , he must be sane, he must have an intact personality, he must have as much information as that intact personality will admit, and he must have been fully informed by miracle.

Let us call this person ' $A+*$.' $A+*$ is just the sort of person whose advice should be relevant to A 's well-being, I say. $A+*$'s advice would span the whole range of A 's personality and will not include desires that stem from a distorted personality. In this way, $A+*$'s advice will not run afoul of the concerns raised in Chapter 3 about the

amount and quality of information. In later chapters I will call $A+^*$ simply $A+$ to keep the language simple.

CHAPTER 5

TURNING THE DESIRE-SATISFACTION ACCOUNT DEVELOPED ABOVE INTO A THEORY OF PRUDENTIAL WELL-BEING

Thus far, in developing my preferred view, I have not taken into account to any great degree conditions of the narrow concept of well-being discussed in the first chapter of this dissertation. For all that I have said so far, the desire-satisfaction account I have been discussing could be viewed as a theory of value, broadly speaking, rather than as a theory of well-being. Please recall that in the first chapter I say that one's prudential well-being is constituted by one's non-moral, non-aesthetic, and self-regarding interests.

As I will make clear in the subsequent parts of this chapter, the requirement that one's interest be self-regarding if satisfying it is to enhance one's well-being is especially troubling for desire-satisfaction theories of well-being. Desire-satisfaction accounts, including the version I have developed in this dissertation, seem to treat all of one's interests as relevant to one's well-being. To some extent, I am less concerned with the requirements that an interest be non-moral and non-aesthetic if it is to count as relevant to one's well-being. As I hope to make clear later in this chapter, once desire-satisfaction accounts can make the distinction between self- and non-self-regarding interests, the matter of non-moral and non-aesthetic interests is less pressing. Even so, I will argue that desire-satisfaction accounts have a nearly unique resource in excluding one's non-moral and non-aesthetic interests from being relevant to one's well-being. I will sketch a few ways in which a desire-satisfaction proponent would eliminate non-moral and non-aesthetic interests.

5.1 Self-Regarding Interests

5.1.1 Conceptions of Prudential Well-Being and Self-Regarding Interests

I hold that virtually all conceptions of prudential well-being, whether desire-satisfaction conceptions or not, must restrict the set of possible interests relevant to well-being to those which are self-regarding. This is contrary to what many people who endorse theories other than desire-satisfaction accounts maintain. For example, most proponents of happiness theories do not take themselves to owe an account of how happiness fits the conditions of a theory of well-being.¹ Perhaps they think that because it is the agent who has the happiness, nothing more must be said to make happiness self-regarding. However, several kinds or sources of happiness might not be sufficiently self-regarding. For example, there is happiness that arises vicariously (manifested in the saying “I’m happy for you that. . .”) and happiness that depends solely on the happiness of other people (manifested in the saying “I’m happy if you’re happy”) and so on. These examples do not prove anything definite; all I hope to accomplish with these examples is to convince the reader that there might even be a need to restrict mental state theories to make them theories of prudential well-being. The thought that “Well, if it is the agent’s happiness in question, then that is sufficiently self-regarding” works only as well as it would work if the proponent of the desire-satisfaction theory said “Well, if it is the agent’s desire in question, then that is sufficiently self-regarding.” I do not think either of

¹ Sumner is a notable exception. He says: “We come now to the notion of happiness with which we will be principally concerned, that in which you are (have been) happy or your life is (has been) a happy one. Being happy in this sense means having a certain kind of positive attitude toward your life. . .” (Sumner, pg. 145). Sumner calls this kind of happiness “life satisfaction” and he takes it to be the most important kind for a theory of welfare. Sumner’s construal of the most important kind of happiness looks as though it is primarily self-regarding.

these responses works well.² So the problem really seems to be a problem that all conceptions of well-being must address. Furthermore, my specific version of desire-satisfaction account does not have any special resources to deal with the problem of non-self-regarding interests. One can have desires and give advice of many different sorts. To use the language of the ideal advisor theory, one's advisor could give advice that is not advisee-regarding.³ In what follows in this section of the chapter, I will provide a way of differentiating self-regarding interests and non-self-regarding interests for virtually any desire-satisfaction account of well-being. Then I will explore the distinction as it is relevant for my specific ideal advisor account. Interestingly, if I am right, my proposal works almost only for desire-satisfaction conceptions of well-being.

5.1.2 Self-Regarding Desires: The Problem of Self-Sacrifice

One illustration of why desire-satisfaction accounts of well-being are troubled by the distinction between self- and non-self-regarding desires is the problem of self-sacrifice. Lots of hypothetical cases have been described in the literature on well-being that have proved troublesome for the desire satisfaction account of the narrow concept of well-being. Mark Overvold has explored the issue of self-interest and self-sacrifice in several papers. His papers figure prominently in the debates over the desire-satisfaction conception of well-being.

An example of self-sacrifice would be a case in which someone donates money to a medical group so that it can find a cure to a disease. There are lots of background

² The perfectionist and objective list theories have similar, if not far worse, sorts of problems.

³ I am also assuming that I could not solve the problem simply by requiring of all ideal advisors, that they restrict their advice to that which is advisee-regarding, for we would no doubt still want to know in what that restriction consists.

assumptions that must be in place to ensure that the case involves self-sacrifice, such as the fact that the person is not donating the money because he has the disease himself and wants a cure, that his act is voluntary, etc. On at least one intuitive understanding of the example, when the man gives the money to the group, he is making a sacrifice of his own well-being because the money he gives could be spent otherwise in a way that benefits him.

The reason that examples of self-sacrifice at least seem to cause problems for desire-satisfaction accounts of well-being is that in cases of self-sacrifice, a desire is satisfied and so, according to these accounts of well-being, the agent's well-being should increase in such cases. Yet, self-sacrifice seems to require a decrease in well-being.⁴ So desire-satisfaction accounts cannot explain it (at least in an easy way).

A number of philosophers have recognized the problem of self-sacrifice for desire-satisfaction theories of well-being. Here are a few: Richard Brandt, *A Theory of the Good and the Right* (pg. 328-331), Derek Parfit, *Reasons and Persons* (pg. 493-502), T.M. Scanlon, *What We Owe to Each Other* (pg. 115), Ronald Dworkin, *Sovereign Virtue* (pg. 25-28), and Stephen Darwall, *Welfare and Rational Care* (pg. 25-31). Each of these philosophers has argued that desire-satisfaction accounts of well-being need to explain self-sacrifice. In what follows, I would like to set up the problem of self-sacrifice in a more detailed way. I will then go on to explain how desire-satisfaction theories can explain self-sacrifice in an intuitive way.

⁴ Perhaps one thinks that an agent acts self-sacrificially when she merely *intends* to act in such a way that her well-being decreases. If unsuccessful, her desire will be left unsatisfied. Thus, in these cases of failed action, so the example is supposed to go, she has no increase in her well-being even though she acts self-sacrificially. The central worry of self-sacrifice remains for desire-satisfaction account because in the cases where the agent succeeds, it seems her well-being increases. I will assume for this paper that self-sacrifice requires a decrease in well-being, for simplicity. If I am wrong in this assumption, the problem is a bit more complex in formulation, but essentially the same.

5.1.3 The Problem Restated

The problem of self-sacrifice for desire-satisfaction accounts of well-being can be expressed as a supposedly inconsistent triad. It seems that any two of the following three propositions can be accepted without inconsistency, but not all three together:

1. Genuine self-sacrifice is possible.
2. People are motivated by their desires.
3. Well-being is increased when desires are satisfied.

(1) should be read partly as a thesis of psychology—about what can serve as a motive. (1) is the thesis that people can be motivated to act in a way that is really self-sacrificial, rather than just apparently self-sacrificial. (2) is what I will call “Humeanism.” Humeanism, understood very roughly, is a thesis about motivation. The Humean claims that a desire is necessary, though perhaps not sufficient, for motivation. (3) is just a general and rough expression of a simple desire-satisfaction account of well-being. I will not challenge the first two propositions in this dissertation. I am disinclined to reject them for reasons outside of the scope of this dissertation. Moreover, many people find (1) and (2) attractive so if I could explain self-sacrifice in a way that does not require the rejection of (1) and (2), that would be an enormous benefit to my theory. I therefore take it that (3) must be altered in some way. However, since I endorse a desire-satisfaction account of well-being, (3) must be altered in a subtle way so as to remain true to the original motivations of desire-satisfaction accounts generally. Before I explore ways of altering (3), I would like to discuss a way in which one might think that the triad is not inconsistent.

Someone might think that the triad is consistent. He might say that the desire involved in motivation does not need to be the “strongest” or “weightiest” desire. Any desire-satisfaction theorist, presumably, is going to claim that the satisfaction of some

desires increase well-being more than others. The triad might seem consistent if one allows for cases of motivation by a less “weighty” desire when the satisfaction thwarts the satisfaction of a “weightier” desire, or thwarts the satisfaction of a set of desires which, when added together, are “weightier.” In such cases, there would be a decrease in well-being and all three propositions could be true.⁵

Unfortunately, this account does not succeed in explaining genuine self-sacrifice. Firstly, the decrease in well-being in such cases is only a *net* decrease. The satisfaction of the motivating desire, in such cases, still increases well-being. That does not seem right. Secondly, the account does not vindicate the possibility of robust self-sacrifice because it limits self-sacrifice to cases in which the agent does not act as she *most* wants. People who dedicate their lives to feeding the starving either would not act self-sacrificially or would fail to live much of their lives as they most want. Neither seems right. What I am looking for is an explanation of a robust sort of self-sacrifice. We should build that robust character into what counts as a genuine case of self-sacrifice in (1). So, this easy way to make the triad consistent does not work. In what follows, I explore ways of amending (3) in a way that vindicates a robust sort of self-sacrifice while remaining true to the spirit of desire-satisfaction accounts.

5.1.4 Overvold’s Proposals

Several of the authors mentioned above suggest that desire-satisfaction accounts of well-being must be restricted to include only desires that are in some sense “related to the agent’s own life.”⁶ The phrase “related to the agent’s own life” is very vague (as the

⁵ I would like to thank John Deigh for pointing out this possibility to me in conversation.

⁶ The phrase “related to the agent’s own life” or very similar phrases show up repeatedly in some of the above mentioned philosophers’ writings.

authors under discussion readily admit). So, we need a precise account of the distinction between the desires that are relevantly related to the agent's own life and those that are not. For simplicity of expression, let us return to treating the expression "the relevant desire set" to mean "the set of desires the satisfaction of which increase well-being." If desire-satisfaction accounts restrict the relevant desire set to include only those desires that are related to the agent's own life, then the resulting triad (with (3) changed accordingly) can be made consistent. For the scope of the relevant desire set is reduced in such a way that makes room for genuine and robust cases of self-sacrifice—namely, cases in which an agent is motivated by desires that are not relevantly related to the agent's own life.

Mark Overvold has offered a way to restrict the relevant desire set. Overvold does not explain the problem of self-sacrifice in terms of the inconsistent triad, but his proposed restriction to the relevant desire set, if it were plausible, would in effect revise (3) to make it consistent with propositions (1) and (2). In fact, he offers two accounts; he offers one account, finds it inadequate, and then offers an amendment to the first account. He chooses the term "self-regarding" to identify the class of desires that are relevantly related to the agent's own life. Later I will return to his choice of terminology, but for now, let us just take it as it is.

Overvold offers his first account in "Self-Interest and the Concept of Self-Sacrifice." In that paper, Overvold says that the agent must be an "essential constituent" in the desired outcome in order for the desire to be self-regarding (Overvold 1980, pg. 118). More precisely, he says that the way to restrict the relevant desire set is by adding the following necessary condition: a desire belongs in the desire set only if "the

proposition that the agent exists at [time] t is a logically necessary condition of the proposition asserting that the [desired] feature or outcome obtains at t' " (Overvold 1980, pg. 118). Very roughly, what Overvold seems to say is that if the agent's existence is (logically) necessary for the truth of the time indexed proposition which expresses the object of the desire, then the desire is self-regarding. I understand that I have only given a rough construal of what Overvold means, but I think further clarification will be achieved by turning to the problems Overvold and others see for his initial account.

Overvold has been criticized from two sides; one side says that his requirement excludes too much and the other side says that his requirement does not exclude enough. That the necessary condition excludes *too much* has been argued by Richard Brandt. He claims that desires for posthumous fame count as self-interested and yet fail to meet Overvold's condition of agent-existence at the time of desire-satisfaction. (Brandt, pg. 330). People probably do not have clear intuitions about Brandt's example because it seems to conflate two issues in desire-satisfaction theories: (a) whether events that happen after one dies can alter one's well-being when one was living and (b) whether desires for fame when alive count as self-regarding.

Whether events that happen after one dies can alter one's well-being when one was alive is not the issue of this chapter, though it is an interesting topic of discussion I will pursue in the final chapter of this dissertation. I will briefly discuss the issue because Overvold's proposed restriction to the relevant desire set, as it stands, excludes posthumous events from altering well-being. Moreover, Overvold seems to recognize this feature of his proposed necessary condition and thinks it an advantage of his account.

Overvold says that “it is hard to see how anything which happens after one no longer exists can contribute to one’s self-interest” (Overvold, 1980, pg. 108).

Some people have intuitions that conflict with Overvold on the matter of whether posthumous events could alter the well-being one had when one was alive. Rather than enter into the debate about this in the present chapter, instead, I will explore a way of modifying Overvold’s proposed necessary condition to make it compatible with the view that events that happen after one has died might alter the well-being one had when one was living.

Brad Hooker, in his “A Breakthrough in the Desire Theory of Welfare,” has proposed a way of modifying Overvold’s proposed necessary condition to allow for posthumous events to alter the well-being one had when one was alive. Hooker proposes: “We might modify this [Overvold’s proposed necessary condition] so that the relevant desires are the ones in whose propositional content the agent is an essential constituent in the sense that the state of affairs *is desired under a description that makes essential reference to the agent*” (Hooker, pg. 212, emphasis his). What Hooker seems to have in mind is that if an agent desires a particular state of affairs to obtain and the proposition describing the desired state of affairs includes essential reference to the desiring agent, even if her existence at t is not a logically necessary condition to the desired outcome at t , then the desire is relevant to her well-being. Hooker’s proposal might be too broad—it might allow desires that are, intuitively, not relevant to one’s well-being to count as relevant to one’s well-being. I will not critically explore Hooker’s proposal. However, I do think that either Hooker’s suggestion or something along those lines could modify Overvold’s proposed necessary condition so that it allows some

posthumous events to alter one's well-being. Overvold's formulation of the necessary condition, strictly speaking, does not allow room for posthumous events to alter one's well-being, but it could be modified to allow for such cases. I think that the spirit of Overvold's proposal is compatible with the view that posthumous events can alter one's well-being. I will not explore the issue of posthumous events now because it is not the central issue of this chapter.

Let us turn now to issue, (b) above—the second issue that is brought to light by Brandt's example of the desire for posthumous fame. The issue is whether desires for fame count as self-regarding. People have conflicting intuitions about this. Suppose that an agent *A* desires that he be famous when alive. Overvold would say that in this case, *A*'s desire is self-regarding. There is essential reference to the desiring agent because *A*'s desire is that *he* be famous. Perhaps the case is controversial in that our intuitions conflict on whether to classify *A*'s desire as self-regarding or non-self-regarding. But note that in this case, there is no issue about Overvold's condition excluding too much, because Overvold's account classifies *A*'s desire as self-regarding.

I will briefly sum up this discussion. I think that neither Brandt's case of desire for posthumous fame nor the revised case of a desire for fame when alive is a clear counterexample to Overvold's proposed necessary condition. The issue of whether posthumous events might alter the well-being one had when one was alive is a separate issue and I think Overvold's proposed necessary condition can be modified to account for it. The issue of whether the desire for fame when alive is self-regarding might be controversial. Overvold's proposed necessary condition classifies the desire as self-regarding. If there is some controversy about the classification, the worry would be that

Overvold's account includes too many desires, rather than that it excludes too many desires.

Thomas Carson, Overvold himself, and others, have argued that Overvold's initial proposal does *not exclude enough*. Carson argues that certain desires, such as desires to accomplish things, will satisfy Overvold's proposed necessary condition and yet should be classified as non-self-regarding (Carson, pg. 236). For example, someone can desire that she secure the money to fix up a run-down park. She is an essential constituent in the desired outcome, so the desire clearly meets the proposed condition about the person being an essential constituent and yet it might seem that it is not self-regarding. A rough way of stating this first sort of problem is that it turns on desires that have the general form: *A desires that A does X*.

Overvold recognizes a different sort of case where his own initial theory does not exclude certain desires that should be excluded from being self-regarding. In a 1982 paper titled "Self-Interest and Getting What You Want," Overvold says that a case involving *A*'s desire that his present wife is happy, and similar cases, cause problems for his initial proposal (Overvold 1982, pg. 189). For, at least in some contexts, it seems that someone's desire that his wife be happy is non-self-regarding even though he must exist in order to have a wife who is happy. A rough way of stating the second sort of problem is that it turns on desires with the general form: *A desires that his Y is Z*.

There is a third very similar sort of problematic case. Imagine a case where *A* desires that the man sitting next to him is cured. *A*'s existence is necessary for the desired outcome, it seems, because *A* must exist in order for a man next to him to be cured. People will have conflicting intuitions, but it seems that in at least in some

contexts, such a desire would be non-self-regarding and yet Overvold's initial proposal would apparently count all such desires as self-regarding. One way of understanding the desires under discussion is by noting they have the general form: *A* desires that *Y*, which bears relation *R* to *A*, is *Z*. This general formulation is to be understood as only very roughly capturing the third sort of problem cases.

Overvold, in his 1982 paper "Self-Interest and Getting What You Want", proposes an amended necessary condition, which is meant to be a restriction on his first proposed necessary condition. Here is what he says.

To handle this objection, let us introduce the notion of the *reason* that a person desires that a particular state of affairs obtain . . . For the desire to be relevant to one's welfare, the reason that one wants the state of affairs to obtain must be due to one's essential involvement in that state of affairs. Consider, for example, *S*'s desire that he bring it about that his wife is happy. If the only reason for this desire is an independent desire that the person who happens to be his wife be happy, then the desire does not seem to be logically relevant to the determination of his welfare. If he lacked an independent desire for his wife's happiness, his essential involvement in the more complex state of affairs would not give him any motivation to perform an act that would have this outcome. For this reason, the desire should be excluded from the determination of his self-interest. (Overvold 1982 pg. 189-190)

Overvold's explanation is a bit confusing because he uses an example that combines two types of problematic cases: it involves a case of a desire that can be characterized as having the general form of a desire by *A* that *A* do *X* (problem case of the first type discussed above) and it involves a case of a desire that can be stated as having the general form of a desire by *A* that his *Y* is *Z* (the second type problematic case discussed above). Let us restrict our consideration to a desire by *A* that his, *A*'s, wife is happy. A bit later in this chapter, I will argue that cases of the first sort are not problematic once we get clear on an independent restriction to desire-satisfaction theories.

The new requirement seems to be that the *reason* for the desire be self-regarding. Unfortunately, what it is to be a *reason for a desire* is not clear. Indeed, just what it is to be a *reason* is not clear. One thing that is clear is that Overvold thinks that his new proposal is an amendment to his first proposal. The final result seems to be that a desire is self-regarding only when the proposition expressing the object of the desire makes essential reference to the desiring agent and when the reason for the desire is due to one's essential involvement in the state of affairs.

At one point in the quotation immediately above, it looks as though Overvold is considering cases in which *A* desires that that woman is happy and also believes that that woman is his wife, and so has a derived desire that his wife is happy.⁷ In such cases, on this interpretation, where no other desires are considered, the desire that his wife is happy should not count as self-regarding because the desire is derived from a non-self-regarding desire. Let us call the interpretation the “derived desire” view. The derived desire view is that a self-regarding desire is a desire for a state of affairs where the state of affairs involves the desiring agent in an essential way and where the desire is not derived from a further desire for a state of affairs that does not involve the desiring agent. Overvold writes of there being an “independent desire,” which seems to support the derived desire interpretation. Also in support of the derived desire view is the fact that it clarifies, at least to some extent, what a person's reason for a desire could be. One simple way of understanding a person's reason for a desire, *D*, is as a desire/belief combination or merely a desire that the person uses to derive the desire *D*. So there seems to be some

⁷ Overvold might also be considering cases in which one desires that that woman is happy and sees no other way for her to be happy than to help, and so derives the desire that he make her happy. This alternative reconstruction is along the lines of problem cases of the first sort. I will deal with problem cases of the first sort later in this essay.

evidence that Overvold has in mind the derived desire view of a reason for a desire. In the part of my chapter that follows, I will argue that Overvold cannot have in mind the “derived desire” view. There is something right about Overvold’s second proposal, but it is important to see why the derived desire interpretation cannot work.

5.1.5 A Problem with the Derived-Desire Interpretation of Overvold’s Second Proposal

The interpretation of Overvold’s second proposal as involving derived desires fails because it includes too many desires as self-regarding, but for an unexpected reason. Desire-satisfaction accounts should include only intrinsic desires as relevant to well-being. An intrinsic desire, roughly, is a desire for some state of affairs in itself. A necessary condition of a desire for x being intrinsic is, roughly, that one not desire x because one believes that it will lead to some further desired state of affairs.⁸ An extrinsic desire, roughly, is a desire for something, x , that the agent has because the agent believes x will lead some further good (the satisfaction of some other desire).

If extrinsic desires were relevant to well-being, then, when both intrinsic desires and the extrinsic desires derived from them were satisfied, there would be a problem of double-counting for well-being. If A intrinsically desires to drink water and for this reason desires to drink from the faucet, then if the latter desire is satisfied, then his intrinsic desire will be satisfied as well.⁹ There are other sorts of extrinsic desires, but the

⁸ Of course, one may have an intrinsic desire and *believe* that satisfying it will lead to some further state of affairs, but one may not have an intrinsic desire *because* one believes that satisfying it will lead to some further state of affairs.

⁹ I understand that my example of an extrinsic desire is of an instrumental desire, which is not exactly the same as Overvold’s example, which appears to be what we might call a “constitutive” desire. A desire is constitutive if the desired thing is desired because the agent believes that having it will constitute satisfying another desire. The belief that that woman is my wife does not link up the two desires instrumentally, as though the satisfaction of the desire that that woman is happy is a means to the satisfaction of the desire that his wife is happy. Instrumental desires are easier to understand and so I will stick with them.

main point is that if extrinsic desires were included in the relevant desire set along with intrinsic desires, then the result would be double-counting when they all are satisfied.

Double-counting, I argue in Chapter 3, is problematic for several reasons. Let me rehearse those arguments here for the reader. For one, double-counting makes one's well-being depend to a large extent on the number of extrinsic desires one has, and this seems implausible. For another, endorsing double-counting would seem to recommend giving priority to satisfying intrinsic desires that are difficult to satisfy or satisfying desires in complex ways, because, in part, if an intrinsic desire is difficult to satisfy, an agent will have to derive and satisfy more extrinsic desires to bring about the intrinsically desired state of affairs. Let me illustrate.

For example, imagine that one has an intrinsic desire to drink water and there is one glass of water front of him and another glass of water beside it locked in a box. On the derived-desire view, in this example (and excluding other intrinsic desires one might have), the agent would benefit more from unlocking the box to get to the water because she will satisfy many extrinsic desires by unlocking the box. The view that allows double-counting is implausible.

We still have yet to get clear on what a reason for a desire is. I certainly do not mean to say that it does not make sense to think of a reason for a desire as being a desire/belief combination or merely a desire in certain contexts. But we need a way of understanding a reason for a desire that can be incorporated into Overvold's proposal without giving rise to the double-counting problem.

5.1.6 Self-Regarding Desires

In light of the above discussion, think that there are five test-cases that a precise account of the distinction between desires that are relevantly related to the agent's own

life and those that are not relevantly related to the agent's own life must be able to handle appropriately.

Agent *A* desires that:

- (a) this stranger is cured of his illness
- (b) the man sitting next to me [*A*] is cured
- (c) my [*A*'s] wife is cured
- (d) I [*A*] cure this stranger
- (e) I [*A*] am cured

Now, the desire for (a) is clearly not relevantly related to *A*'s life and the desire for (e) clearly is relevantly related to *A*'s life. The cases of desires for (b), (c), and (d) are perhaps controversial. A clear account of the distinction must be able to explain desires for (a) – (e) in an intuitive way.

Carson thinks that a desire for (d) is a problem case for Overvold's account and Overvold himself seems to think so as well. I think that, once we restrict the relevant desire set to intrinsic desires, desires that *p* where *p* is of the form "I do *X*" turn out to be intuitively self-regarding. If a woman has a desire that a park be repaired and, because no one else is willing to do the work, she derives the desire that *she* raise the money, then it is plausible to think that the satisfaction of her derived desire would not enhance her well-being. However, if a woman has an intrinsic desire that *she* raise money for a run-down park, then excluding other desires, she would be *disappointed* if she found out that someone else raised the money. This intrinsic desire is self-regarding in the relevant way and so its satisfaction should count as enhancing her well-being.¹⁰ So desires that (d) and relevantly similar cases are not problematic. If it is an intrinsic desire, then it belongs in the relevant desire set.

¹⁰ I will return to Overvold's choice of "self-regarding" to capture the notion that a desire must be relevantly related to the agent's own life in order for it to be relevant to her well-being later in this chapter.

The cases of (b) and (c) are a different matter. To deal with these cases, I think it helps to look at two ways of construing a desire: *de dicto* and *de re*.¹¹ A desire is *de dicto*, it is said somewhat vaguely, if it is “about the words” and *de re* if it is “about the thing.” Additionally, it is said that that *de re* attitudes relate a person to an object while *de dicto* attitudes relate a person to a proposition. Perhaps it is best to illustrate this difference by an example. One can have a *de dicto* desire that the man with the bowler hat is cured. Alternatively, one can have a *de re* desire, of the man with the bowler hat, that he is cured. The *de dicto* desire under discussion is for a certain state of affairs to obtain. One can have this desire even if there is no man in a bowler hat. The *de re* desire under discussion is more particularly a desire, with respect to an object, that *it* have some characteristic. One cannot have such a desire unless the relevant object exists. If there is no man in a bowler hat, one cannot desire, of the man in the bowler hat, that he be cured.

Let us see whether the distinction between a *de re* and *de dicto* desires, so understood, gets us any clarity with the remaining problem cases. Take my desire that the man sitting next to me is cured (desire type (b) above). It could be a *de re* desire in which case it would be a desire, of the man sitting next to me, that he is cured. In such a case the reference to me is clearly just to enable me to pick out the man.¹² The desire is not about me. This case is very similar to my desire, of the man with the bowler hat, that

¹¹ Philosophers have provided different ways of understanding *de re* and *de dicto* desires. My characterization of the types of desire is quite different from some of them. I will not explore the various ways in which philosophers have characterized the distinction, but I hope to make clear at least one way of characterizing the distinction in this essay.

¹² I understand that desires are not sentences and sentences are not desires. It is sentences that express desires and it is sentences that “make” reference to the agent. I will adopt a more careful way of speaking when I arrive at a more complete account of the distinction between self-regarding and non-self-regarding desires. For simplicity of expression at the present, I will write as though desires, in some cases, make reference to agents.

he is cured. Indeed, if the man next to me *is* the man in the bowler hat, then the desires are one and the same. Such desires, intuitively, are non-self-regarding.

Alternatively, my desire could be *de dicto*. My desire could be the desire that the man sitting next to me is cured. In such a case, where the desire is intrinsic and *de dicto*, and is not better characterized as *de re*, “of the man,” it is self-regarding. Let me explain this. I could very well have a phobia against being near sick people. Such a phobia could lead to an intrinsic desire that the man sitting next to me is cured. My phobia would not be my reason for the desire, but would be part of an explanation of the desire. In this case, my *de dicto* desire that the man sitting next to me is cured would be relevant to my well-being. My desire is relevantly related to my life.

The distinction between *de dicto* and *de re* desires can also explain how desires of type (c) should be treated. If A’s desire is *de re*, then A desires, of the woman who is his wife, that she is cured. In such a case, the reference to A is only there to pick out the object of his desire and the desire is non-self-regarding. The *de re* reading would treat A’s desire as very similar to a desire, of the woman in the blue dress, that she is cured. Indeed there might only be the one desire if A’s wife is the woman in the blue dress. In both cases, the desire is not self-regarding and so its satisfaction does not enhance the agent’s well-being. I admit it is hard to imagine a context in which A’s desire that his wife be cured—his desire in (c)—is *de re*. Perhaps we can imagine a case in which A’s desire, of his wife, that she be cured, stems from A’s caring for his wife for her sake. Perhaps A’s love of his wife influences him to, or consists partly in wanting to, care for her for her sake. If A’s desire stemmed from or involves his concern for her for her sake, then his desire could plausibly be thought to be *de re*. In any case, I think we can see that

A's desire for his wife to be cured would, in some context, be *de re*, and, in other contexts, be *de dicto*.

There is a line in the above quotation from Overvold that suggests he has in mind something along the lines of the *de re/de dicto* distinction. Overvold says "If the only reason for this desire is an independent desire that the person who happens to be his wife be happy, then the desire does not seem to be logically relevant to the determination of his welfare" (Overvold, 1982, pg. 189-190). Overvold writes of A desiring that the person who *happens* to be his wife be cured. It seems obvious that Overvold does not have in mind that the man is married to someone by some strange coincidence. Instead, Overvold might have in mind that the man picks out the person who is his wife with a possessive pronoun, but wants her to be happy independently of her being his wife. If that is what Overvold has in mind, then the desire is *de re*. Perhaps, then, the *de dicto/de re* distinction is a clear way to capture his thought.

Let me explain the *de dicto/de re* distinction in a technically correct, though more abstract way. Some desires, though they may at first glance seem like *de dicto* desires, are more accurately construed as *de re*. Let me introduce a bit of terminology. Let us say that the "subject clause" of a desire attribution is the part of the sentence that ascribes the desire that does not express the object of the desire, but that instead contains the reference to the desiring agent, possibly the object picked out, and "that" or something synonymous. For example, the "subject clause" of the sentence "A desires that the person sitting next to him get well" is "A desires that." In this example, note the reference to A. In the subject clause of desire attributions, there will be some reference to the desiring agent, because the desire must be attributed to someone. Now, in some subject clauses,

there will be a *second* reference to the desiring agent. For example, in the “subject clause” of “A desires, of the person sitting next to him, that he be well” there are two references to A. The first reference to A just attributes the desire to A, while the second reference to A helps pick out the *de re* object of desire. The first reference to A will not make the desire the sentence is ascribing be self-regarding. A second reference to A will not make the desire self-regarding either.

Let us return to the sentence “A desires that the person sitting next to him get well.” The part of the sentence that reads “the person sitting next to him get well” we can call the “object sentence.” A sentence that expresses a *de re* desire will also have an “object sentence.” In the sentence “A desires, of the man sitting next to him, that he be cured,” the “object” sentence is “he be cured.” Some *de re* desire attributions will make reference to the desiring agent in referring to the object of desire in the “subject clause,” and they may ascribe a non-self regarding desire. For example, if A desires, of the man sitting next to him, that he is well, then A’s desire is non-self-regarding. The second reference to the agent in the subject clause is merely there to pick out the *de re* object i.e., maybe, the man in the bowler hat. Many desires will be *de dicto*. Desire attributions that make reference to the desiring agent in the object sentence will ascribe self-regarding desires.

There is one further complication. There may be cases of *de re* desires where the agent, herself, is the *de re* object of the desire. As a hypothetical example, it seems entirely possible that I could desire, of myself, that I be cured of my disease. There is the usual reference to me in the “subject clause” (merely attributing the desire to me), a *de re* reference to me in the “subject clause” (helping to pick out the *de re* object of desire), and

a reference to me in the “object sentence.” In such a case, the desire is self-regarding. Nothing in my treatment of the case under discussion is incompatible with what I say above. Let me explain. One *might* think that the *de re* reference to me in the “subject clause” makes the desire non-self-regarding, but this would be a mistake. The reference to me in the “object sentence” ensures that the desire is self-regarding. The lesson we can learn from the example of my desire, of me, that I be cured, is that one cannot look just to the “subject clause” of the sentence which expresses the desire in trying to decide whether a desire is self-regarding or not. One must look for reference to the desiring agent in the “object sentence” as well. If one has a desire such that any sentence that aptly expresses the desire makes reference to the agent in the “object sentence,” then the desire is self-regarding.

Here is the general principle that makes the distinction between the desires that are self-regarding and those that are not: A desire is self-regarding if and only if (a) it is *de dicto* and any sentence that aptly expresses the desire makes reference to the desiring agent in the “object sentence,” or (b) it is *de re* and any sentence that aptly expresses the desire has reference to the desiring agent in the “object sentence.” In both *de dicto* and *de re* desires, there has to be reference to the desiring agent in the “object sentence.” Reference to the agent in the “subject clause” of a sentence expressing the desire does not make a desire self-regarding. This seems a clear way of drawing the distinction and is intuitive.

Let me briefly revisit the problem cases above. Take A’s desire that the man sitting next to him [A] is cured (case (b) above). Desires of this sort will be self-regarding in some contexts and non-self-regarding in others. If A’s desire in this case is *de re*, then

the desire is non-self-regarding, but the desire is self-regarding if it is *de dicto*. Take the case of A's desire that my [A's] wife is cured (case (c) above). If the desire is *de re*, then it is non-self-regarding, but it is self-regarding if *de dicto*.

I am not certain that Overvold had in mind exactly what I have proposed. It may even be that Overvold was right in his first proposal if it is properly interpreted. For, perhaps one way of understanding just what it is to be an essential constituent in a desire is on the *de re/de dicto* model. One further thing to bear in mind is that not all self-regarding desires will be relevant to one's well-being. Only desires that are both intrinsic and self-regarding are relevant to one's well-being.

There may be a bit of dissatisfaction with my account. One may still find it implausible that some desires that my account treats as self-regarding are genuinely self-regarding. Take for example A's desire that she help the poor. Even if it is intrinsic, one might say that intuitively it is non-self-regarding. I think that any such intuitions to this effect might be caused by Overvold's choice of terminology. He chooses to use "self-regarding" to identify the key class of desires to which we must restrict desires that are included in the set of desires the satisfaction of which increase well-being (the so-called relevant set). I mentioned above that some authors say that in order to solve the problem of self-sacrifice for desire-satisfaction accounts of well-being, desire-satisfaction theorists must restrict the relevant set of desires to those that are "related to the agent's own life." Instead of "self-regarding," perhaps it would be better to use "self-involving." "Self-involving" seems to have a broader scope and self-sacrifice may be better explained in terms of self-involving desires than self-regarding desires. Let me explain.

When one speaks of “self-regarding” desires, the reader may get the impression that such desires have to be selfish in some way. But it would be a mistake to understand well-being or self-interest to be a matter of pursuing selfish wants. Selfish desires are those desires that one has that, if satisfied, enhance one’s well-being to the exclusion of the well-being of others. In my opinion, no restriction to selfish desires should be placed on the relevant desire set for an account of well-being or self-interest. One can have desires that are related to the agent’s own life that also end up benefiting other people. So *A*’s desire that help the poor may very well end up benefiting other people in such a way that it is not selfish, but it is related to the agent’s own life and is self-involving. I do not know whether speaking of “self-regarding desires” suggests that the desires must be selfish. But perhaps it does. If Overvold had used “self-involving,” any potential confusion on this point could have been eliminated at an earlier stage. In any event, I will continue to speak of “self-regarding” desires, but it is important to understand that such desires need not be selfish.

5.1.7 Self-Regarding Interests Concluded

I take myself to have given an adequately precise account of the distinction between self-regarding and non-self-regarding desires. I feel that I have provided an adequate explanation as to how the desires in (a)-(e) should be treated. (a) clearly involves a non-self-regarding desire. (e) clearly involves a self-regarding desire. (d) intuitively involves a self-regarding desire once we understand that any desire must be intrinsic to be relevant to well-being. The desires in (b) and (c) are non-self-regarding, if we assume that they are *de re* and that there is no reference to the desiring agent in the “object sentence” of the sentence expressing the desire, and they are self-regarding otherwise.

Once the distinction between self- and non-self-regarding desires is in place, self-sacrifice can be explained in a manner that is consistent with the spirit of desire-satisfaction theories of well-being. If the desires, the satisfaction of which contribute to enhancing one's well-being, are restricted to desires that are self-regarding, then an act of self-sacrifice will occur when an agent is moved by a non-self-regarding desire the satisfaction of which thwarts the satisfaction of a self-regarding desire.

Let me return to the triad mentioned above. The first proposition, (1), is that genuine self-sacrifice is possible. The second proposition, (2), is that people are motivated by their desires. The third proposition, (3), is that well-being is increased when desires are satisfied. The triad of propositions is inconsistent. Desire-satisfaction theories must be modified and restricted slightly. If (3) is amended to reduce the scope of the relevant desire set, then the resulting trio of propositions is consistent. There can be genuine acts of self-sacrifice. A desire is necessary to motivate action (even if desires are understood so broadly as to include any mental state that motivates). Finally, the satisfaction of self-regarding desires increases well-being. I hope to have dealt with a problem for desire-satisfaction theories in a clear and intuitively attractive way.

I need to explain how my ideal advisor account will deal with the distinction between self- and non-self-regarding desires. My ideal advisor account, very roughly, is that one's well-being varies with the extent with which the desires one's fully informed counterpart would have for one are satisfied. The way I envision my ideal advisor account, one's ideal advisor, when contemplating his relatively less informed counterpart and thinking about what advice to give, would have a list of desires the satisfaction of which would make one's life go better. If A^+ is A 's ideal advisor, A^+ would want many

things for A. A+'s desire is *for* A in the relevant way, I say, if and only if A+'s desire is *A-regarding* in the technical way discussed the present section. That is to say, A+'s desire is for A, in the relevant way, if it makes reference to A in any "object sentence" aptly expressing A+'s desire, and not merely in a *de re* "subject clause."

Let me illustrate with an example. Imagine that A+ has five intrinsic desires. In particular, A+ desires: (1) that A+ win the Boston Marathon, (2) that A read *The Metamorphoses*, (3) that A do his moral duty to help the poor, and (4) that A+ eat a bowl of ice cream and (5) that A eat a bowl of ice cream. Now, on my account, only desires (2), (3) and (5) will be A-regarding and so only (2), (3) and (5) will be relevant to A's well-being. Desires that have as their "object sentences" (1) and (4) are non-A-regarding and so are not relevant to A's well-being.

At this point, I would like to move onto the next area where my account of well-being developed in chapters 3 and 4 can be amended to fit the conditions of prudential well-being discussed in Chapter 1.

5.2 Non-Moral and Non-Aesthetic Interests

5.2.1 The Problem

In the first chapter of this dissertation, I argued that one's prudential well-being is a matter of satisfying one's non-moral, non-aesthetic, and self-regarding interests. Thus far in this chapter, I have discussed how one can make desire-satisfaction accounts of well-being fit the "self-regarding condition" of prudential well-being. It is possible, even with the relevant desire set restricted to self-regarding desires, to have a desire that is moral in the relevant way. We have already seen an example for my ideal advisor theory, but let me provide an example that applies to all desire-satisfaction account. A person could have a desire that she do her duty. Her desire is that she do *her* duty, in my example.

Her desire, let us say in this case, is self-regarding, because any sentence aptly attributing the desire to her makes reference to her in the “object sentence.” In such a case, the desire seems self-regarding and yet the desire seems to have moral content in that any sentence aptly expressing it makes reference to the moral concept of duty in the relevant way.¹³

My goal of excluding one’s moral interests from being relevant to one’s prudential well-being might strike the reader as surprising. To the reader who finds my goal surprising, I have two comments. First, I think that my examples from the first chapter show, at least in an intuitive way, that moral, aesthetic and prudential values can be separated. My examples from *The Count of Monte Cristo* were meant to show that an immoral course of action can be in one’s prudential interests nevertheless. Think back to the example of Monte Cristo trying deceptively to marry his adversary’s daughter and her half-brother. Surely there is something immoral about the project and yet it seems intuitive that success in immoral projects can make one’s life go better, at least in a prudential way. Second, a number of philosophers make reference to the individual good or the non-moral good of a person. One notable example is Peter Railton who claims to be giving an account of the non-moral good of a person in his essays “Moral Realism” and “Facts and Values.” Railton, who favors an ideal advisor account of sorts, and others who write about the non-moral good of a person, leave open the way with which they would deal with desires that have moral content.

¹³ If the concept of duty, surprisingly, turns out not to be a (coherent) moral concept, then I feel that another example of the relevant sort could be given. The problem-case that I have in mind need only make use of one moral concept, whichever that concept is among the usual candidates.

If I am right, then all conceptions of the concept of prudential well-being must deal with variations of the problem I discuss above. For example, one could be happy or pleased that one has done one's duty. Any mental state account of prudential well-being that treats one's happiness or pleasure as constitutive of one's well-being must deal with the case in some way. Proponents of such accounts have not dealt with problem cases of the sort under discussion. Objective accounts of well-being, such as accounts that treat one's well-being as varying with the extent to which one exercises and develops one's nature, must also deal with problem cases of the sort under discussion. There is not anything in the objective human nature accounts, as they are proposed, to deal with such cases.

It is not clear how proponents of the various conceptions of well-being would deal with the problem cases and ensure that their accounts distinguish a person's prudential good from her moral good. There is not much explicit discussion of this problem in the literature. It seems to me that proponents of happiness conceptions would be willing to "bite the bullet."¹⁴ They accept that one can be happy, say, that one has done one's duty, and are even willing to accept that happiness that one has done one's duty increases one's well-being, at least to some extent. Likewise, proponents of human nature conceptions of well-being might accept that it is or might be in one's nature to act in a moral way. The proponent of human nature conceptions might also accept that exercising and developing one's nature in a moral direction increases one's well-being. Desire-satisfaction theorists

¹⁴ Given that happiness proponents are willing to grant that happiness that is in a relevant way non-self-regarding nevertheless constitutes some well-being, I suspect that they are also willing to grant that happiness that is self-regarding but moral constitutes some well-being. Their answer, I predict, is very similar to the one they give to examples of non-self-regarding happiness: to wit, that if it is an agent's happiness, then it constitutes the agent's well-being.

can “bite the bullet” as well on this issue. A desire-satisfaction proponent could claim that the desire that one does one’s duty is in the relevant desire set. However, I think that desire-satisfaction theorists can do better, as I will explain in the following parts of 5.2.

5.2.2 Moral Desires and the Relevant Desire Set

Earlier in this chapter I wrote of the so-called relevant desire set, which is the set of desires the satisfaction of which increase well-being. A proponent of desire-satisfaction accounts can restrict the relevant desire set to those desires that are non-moral. Let me return to an earlier example. Suppose an agent, *A*, has an intrinsic desire that she, *A*, do her duty. Her desire, in my example, is self-regarding. Yet her desire, so described, also has moral content. *A*’s desire, in this case, is one that a desire-satisfaction theorist might wish to exclude from the relevant desire set given that the desire has moral content and given that the desire-satisfaction theories under consideration, let us stipulate, are theories of prudential well-being.

That is, a desire-satisfaction theorist, in her attempt to exclude moral interests from being relevant to well-being, could adopt a strategy very much like the one she can adopt in her attempt to exclude non-self-regarding interests from being relevant to well-being—simply restrict the relevant desire set.

I think of my ideal advisor account as being able to be restricted in the ways indicated in this chapter thus far. There could be a problem however. As I mention above, on the way I envision my ideal advisor account, one’s ideal advisor, when contemplating his relatively less informed counterpart and thinking about what advice to give, would have a list of desires the satisfaction of which would make one’s life go better. If we again speak of *A*⁺ as *A*’s ideal advisor, *A*⁺ would want many things for *A*. In the example above, three of *A*⁺’s desires are sufficiently *A*-regarding. They are *A*⁺’s

desires (1) that *A* read *The Metamorphoses*, (2) that *A* do his moral duty to help the poor, and (3) that *A* eat a bowl of ice cream. Now, in this simplistic and idealized example, I think that if we restrict the relevant set to desires that do not have moral content, the result would be that the only desires that are relevant to *A*'s well-being are (1) and (3). Let us assume that *A*+ desires (1) strongly and (3) much less strongly. However, and now we get to the potential problem, if *A*+ were to not think about *A*'s duty, perhaps *A*+ would put desires (1) and (3) in a different order or perhaps *A*+ would desire different things altogether for *A*. If *A*+ were to contemplate only what he wants for *A* apart from *A*+'s beliefs or feelings about *A*'s duties, *A*+ might rank desire (3) ahead of (1).

As I suggested above, I do not think that such a re-ordering could happen, but I do not have an argument to show that it could not happen. Unless I revert to the “bite the bullet” strategy of simply accepting one's moral interests as relevant to one's well-being, which most other proponents of conceptions of well-being seem willing to do, I need another account—a contingency account—in case a re-ordering of the sort I describe above is possible. If such a re-ordering is possible, I think “biting the bullet” on it would be equally as counterintuitive as “biting the bullet” on the initial problem of moral interests.

5.2.3 Another Approach

There is another way that desire-satisfaction theorists can exclude one's moral interests from being relevant to one's well-being. Moreover, I think that my ideal advisor theory has special resources in dealing with moral and aesthetic interests that other versions of desire-satisfaction accounts do not have. One way that the reader can think of the rest of the present section of this chapter is that in it, I try to show a way in which many desire-satisfaction theories and, more specifically, my ideal advisor theory, can do

better than most other conceptions of well-being in dealing with the problems that arise from moral and aesthetic values for conceptions of prudential well-being.

To see how desire-satisfaction proponents might deal with problem cases of the sort discussed above, I would like to turn, perhaps surprisingly, to John Stuart Mill. As I interpret Mill in his *Utilitarianism*, he does not endorse a desire-satisfaction conception of well-being, rather, he endorses a sophisticated kind of pleasure-based mental state account.¹⁵ However, he is concerned that moral values may affect his account of utility. Mill describes a test for determining which of one's pleasures count as increasing one's utility, given a set amount of intensity and duration. The test is popularly known as the "Test of the Competent Judges". Of this test Mill states: "Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure" (Mill, Ch. 2, Par. 5). Mill, as I interpret him, proposes that if all or most "competent judges" decidedly prefer one pleasure to another, given identical intensity and duration between the two pleasures, then the preferred pleasure counts for more utility, so long as the preferences of the competent judges are not influenced by feelings of moral obligation. Mill fairly clearly attempts to exclude moral influence from altering the judges' decided preferences.

As I note above, Mill is not a desire-satisfaction theorist. Perhaps what he calls preferences are what I think of as desires; thus his and my theory might have significant similarities. At most, though, Mill could endorse a mixed conception of well-being. He

¹⁵ For the purposes of this example, it does not matter whether utility and prudential good are the same thing. Though for what it is worth, I think that Mill's concept of utility is very close, if not identical, to the concept of prudential well-being I explicate in the first chapter of this dissertation.

could hold a view that treats both one's mental states and one's desires as relevant to one's well-being. What is important for this section of this chapter, however, is not getting clear on the precise nature of Mill's theory, but rather, a solution to a problem he sees for his account of utility that can be adopted by many desire-satisfaction accounts of well-being.

It is clear that Mill has in mind a counterfactual condition, somewhat similar to the counterfactual condition I set out for my ideal advisor theory. Using "possible worlds" language, Mill's "Test of the Competent Judges" might be revised and explained as follows: The decided preference for one of two pleasures by our counterparts in the nearest possible world who have experienced both pleasures and are capable of appreciated both pleasures is the sole criterion for which is the higher pleasure. Though Mill was writing about higher and lower pleasures, what he says about preferences and moral values can be adapted for use by many desire-satisfaction theorists. The key similarity between Mill's account as I have reconstructed it and many desire-satisfaction theories is that both treat the central criterion for their respective accounts as a counterfactual preferences or desires.

Mill does not describe how the "competent judges" are to think in any detailed way. Two intuitively attractive ways of understanding what Mill has in mind are: (1) the competent judges are our counterparts in the nearest possible world where we (or they) simply have no moral feelings whatsoever, and (2) the competent judges are our counterparts in the nearest possible world in which we (or they) disregard, to the best of their abilities, their moral feelings. Similarly, desire-satisfaction theorists could say that either (1) *A*'s well-being is constituted by the satisfaction of *A*'s desires for *A* in the

nearest possible world in which $A+$ lacks any moral feeling whatsoever, and suitable other conditions, such as a condition of being fully informed, are met, and (2) A 's well-being is constituted by the satisfaction of $A+$'s desires for A in the nearest possible world in which $A+$ disregards A 's moral feelings, to the best of $A+$'s abilities, and suitable other conditions are met.¹⁶ Rather than pursue what Mill might have had in mind with his "Test of the Competent Judges," I would like to very briefly explore the options for counterfactual desire-satisfaction theories.

Option (1) is an intriguing way of excluding one's moral interests from being relevant to one's well-being. But, the nearest possible world in which one has no moral feelings whatsoever is perhaps quite a ways away. One's counterpart in such a world would be incapable of desiring that one's actual self do his duty. Notice that on the proposal under discussion, there is no restriction of the relevant desire set to exclude moral interests, for one's counterpart could not have a desire with moral content. There is no room for the potential problem of re-ordering.

To some extent, I worry that on option (1), the removal of moral feelings at least for some people would result in a different personality. This worries me since one of the central motives for adopting my ideal advisor account is that it is the best account that isolates the desires that are true to one's personality and reflect the full range of one's personality. I am concerned about changes that might alter one's personality. This is somewhat speculative, of course, but I think that the second option is a more promising

¹⁶ I see no harm, on either options (1) or (2), in allowing $A+$ having knowledge of A 's moral feelings and no harm in using information about A 's moral feelings when forming advice for A . A 's moral feelings are not, in themselves, $A+$'s moral feelings. Rather, A 's moral feelings are merely part of the information base that $A+$ uses in formulating advice. It seems that information about A 's moral beliefs, for our purposes here, is similar to any other sort of information $A+$ could use to formulate advice.

way to deal with the problem of moral interest, if there is indeed the potential for a re-ordering of one's advisor's desires in cases of the sort mentioned above.

Option (2), I think, is a very good way of dealing with the sort of problem case under discussion. *Disregarding* one's moral feelings might seem to the reader to be impossible, but I think that it is both possible and done on a regular basis in cases or everyday judgments of well-being. Think back to some of the Monte Cristo cases found in the first chapter of this dissertation. We have moral feelings, I take it, and yet we managed to disregard them when we made judgments about what is good for Monte Cristo. To use a vivid example, think of assassins. Though we feel that their behavior is morally reprehensible, we can easily think of something as good for the assassin without feeling it is morally good for him. Such everyday examples make me think that disregarding our moral feelings in assessing someone's prudential good is not an uncommon thing. Bear in mind as well that a theorist who takes option (2) in dealing with moral feelings is trying to identify the nearest possible world in which one is able to disregard his moral feelings to the greatest extent possible, while holding constant one's personality.

Since my ideal advisor account is supported by considerations involving the personality of a person when assessing what contributes to her well-being, it makes sense to require that the nearest possible world in which she disregards her moral feeling to the greatest extent possible be one in which her personality is the same. This allows for the possibility that one's personality is so constituted that she cannot disregard her moral feelings in giving advice to her counterpart. But I think that this is just as it should be.

If, indeed, a person is so constituted that she cannot disregard her moral feelings, then those feelings are relevant to her prudential well-being, I say.

Before I finish my discussion of moral values and prudential well-being, I will briefly discuss one additional potentially problematic case. We can imagine that a fully informed idealized counterpart is successful in disregarding her moral feelings, but nevertheless, has a virtuous character in the sense that her advice tracks the non-moral features of the world upon which morally good character or morally good actions are based. For example, a person with a virtuous character might often choose to alleviate the suffering of another, even absent influence from moral beliefs. The virtuous person, on this slender account, tracks the good-making features of the world, but not the good itself. Nothing that has been said so far excludes such advice from counting. In fact, I think such advice should count as relevant to a person's well-being, provided that it is sufficiently advisee-regarding. For example, suppose A+'s advice to A is that A alleviate the suffering of another, and suppose that this advice flows from her virtuous character. If so, then I think it plausible to allow it to count as relevant to A's well-being.

5.2.4 Aesthetic Values and Prudential Well-Being

A brief discussion about aesthetic value and prudential well-being is in order. Much of what I have said regarding moral value can be adapted to explain how to deal with aesthetic value.

I will briefly explore a problem case. Suppose that one has aesthetic feelings about life stories. For example, suppose that someone is taken by the aesthetic beauty of tragedies and thinks that the best kind of life (in a sense) is the tragic one. A fully

informed person might advise her less informed self to live a tragic life.¹⁷ Surely this advice must be excluded somehow.¹⁸

I think that aesthetic advice of the sort I have in mind can be excluded from being relevant to well-being in the same ways that moral advice can be excluded. Given the way that I discussed moral value, there are four ways one could go, in my mind: (a) one can “bite the bullet” and claim, somewhat implausibly, that desires with aesthetic content are relevant to well-being, (b) one can claim that the relevant desire set should not have any desires with aesthetic content, (c) one can claim, at least on counterfactual desire accounts, that the relevant counterpart must have no aesthetic values, or (d) one can claim, at least on counterfactual desire accounts, that the counterpart should disregard one’s aesthetic values.¹⁹

¹⁷ Once in conversation with me, Elijah Millgram hypothesized that Oscar Wilde lived his life so that it would be a tragedy. Wilde’s life story can most certainly be viewed as a tragedy, but the claim that Wilde lived his life so that it would be a tragedy is a more significant claim. The claim that he lived his life so that it would be a tragedy is not entirely unfounded: Wilde deliberately opened up the possibility of his being prosecuted for breaking the sodomy laws in England and then, upon being found guilty and sentenced to prison, refused to flee the country though given ample opportunity to do so. There really does not seem to be any evidence that Wilde was civilly disobedient—broke the law but accepted the punishment as a way of showing respect for the law (in the way that Martin Luther King Jr. was when he willingly broke the segregation laws and took the punishment with the goal of showing the injustice of the laws). So there does not seem to be any principled reason why he decided to remain in England and go to prison. Wilde knew of the probable outcome of his provocation of the trial and many people close to him, foreseeing the outcome, advised him against his course of action. So perhaps his ideal adviser, even with full information of the consequences would have advised him to act as he did. The issue is very complicated however, and any further debate as to Wilde’s motives must take place elsewhere.

¹⁸ A more controversial case might be one where the fully informed self advises her less informed self to live a life in accordance with a life story that has a supremely happy ending. The happy ending could either *consist in* the person having a high level of well-being or (more likely) merely *involve* a high level of well-being. Comedies, at least many sorts of comedies, are easily eliminable from the list of advice (could one’s ideal advisor recommend to live as Larry, Moe and Curly do?). The ideal advisor could advise one to live one’s life in accordance with a life story that has a supremely happy ending, I suppose, but if the advice is given for aesthetic reasons, it could not count as relevant to one’s well-being—in the narrow sense of well-being. This would be a case where the ideal advisor gives the right advice, but for the wrong reason.

¹⁹ Recall Darwall’s claim that what enhances someone’s well-being is he or she getting what it is rational for us to want for them were we to care for them. If there are ways in which a life can go morally or aesthetically better, I do not see anything in Darwall’s view to exclude this kind of advice from being offered. I do not have children, but if I did, I would want them to have a high moral character, live

As in the cases of desires with moral content, I think that option (b) is attractive. However, since I do not have an argument that shows that re-ordering is not possible, my “fallback” preference is for (d).

5.3 Perverse Advice

5.3.1 Caring and Well-Being

Thus far I have discussed advice that is influenced by substantive values and advice that is non-self-regarding. There seems to be a third kind of advice that could prove troubling for my version of desire-satisfaction theory. The worry is that there are some instances where the advisor could provide advice that is, by our intuitions, *perverse* and the satisfaction of which should not count towards the advisee’s well-being. Such advice could be self-regarding, in a technical sort of way, but seems to run afoul of the spirit behind the requirement of an interest’s being self-regarding as set out in the first chapter of this dissertation. There are two sorts of cases of perverse advice I will explore, one that is not genuinely problematic for my theory and one that is. The way to get around the genuinely problematic cases, I will argue, is to add the requirement that the ideal advisor *care* for his advisee.

I will begin with the first, merely apparently problematic sort of case. I will revert back to my terminology of saying that *A* is some agent as he actually is and that *A+* is his fully informed counterpart who also meets the conditions set out in chapters 3 and 4, and

beautiful lives, and also have a high level of prudential value. I imagine that is what just about everyone wants for his or her children. Darwall, of course, has the restriction that the “wants” must be rational and they must be “for the person,” so perhaps in a fully explicated account of his view, he could exclude moral and aesthetic advice. My suspicion though, is that Darwall’s requirement that the desires be rational does not exclude invading substantive values and his requirement that the desires be “for the person” only excluded instrumental desires; such as the desire that my brother win the lottery *only* because I think that he will give me some of the winnings. Please see my Chapter 1 and section 5.3.2 of the present chapter for further discussion of Darwall.

sections 5.1 and 5.2 of the present chapter. Imagine that $A+$, after having been informed by miracle, now, hates his actual self, say, because $A+$ finds A ignorant. If this were the case, so the objection goes, then $A+$ could recommend to A that he jump off a cliff. Now, I do not think that this sort of case is problematic. The reader might recognize it as a counter-intuitive consequent change similar to the ones discussed in the previous chapter. Obviously, it is not logically necessary for someone who becomes fully informed to hate his less informed counterpart. Recall that we are looking for the nearest possible world to the actual world in which A is fully informed. So long as A does not hate himself, so my thought goes, $A+$ will not hate A . So the first sort of case is not really problematic.

However, my opponent might reply, what of cases in which A hates himself? In the nearest possible world in which A is fully informed, $A+$ will hate A , so the counter-argument goes. Here we have arrived at the second sort of case alluded to above which is a genuinely problematic sort of case for my account of well-being. $A+$'s advice must be A -regarding in a robust way. $A+$'s advice can be A -regarding in the technical way discussed in the previous section of this chapter, but it might not be A -regarding in a more robust sense. My solution, as I mention above, is to require that $A+$ care for A , whether A cares for himself or not. If $A+$ were to care for A , then $A+$'s advice for A would be A -regarding in a robust way. I will explain how the requirement that $A+$ cares for A helps to eliminate perverse advice later, but I wish to deal with concerns that the very idea of the requirement of caring, in this context, is somehow conceptually misguided.

What if caring for someone is identical to wishing that he or she have a high level of well-being? If that were the case, then the account of well-being I am proposing

would be circular. If we try to add the requirement that $A+$ care for A to solve the problem of perverse advice, but if caring is simply wishing that someone fare well, then the account could not get off the ground. So the challenge then, at this stage, is to explain what caring is and how it functions in an ideal advisor account, without explaining caring as involving well-wishing.

In explaining what caring is, and how it is to function in the mind of the ideal advisor, great pains must be taken to ensure that caring is not the same as wishing that the advisee would have a high level of well-being (from now on, I will call this “well-wishing”). Strictly speaking, there are obvious cases of well-wishing that are not cases of caring. Here is one example; say I wish that my brother has a high level of well-being merely in the belief that he would help me out. It would be very misleading to say that I then cared for him because I wished him well. Another case where the two clearly come apart is a case in which one desires to benefit others out of a Kantian motive of moral duty. So well-wishing and caring are not identical, but they still may be close enough to make the proposed solution circular.

5.3.2 Darwall on Caring and Welfare

Darwall develops what he calls a meta-ethical account of welfare according to which something contributes to someone’s welfare if and only if it is something it is rational to want for that person, insofar as one cares for that person (Darwall, pg. 4).²⁰

The penultimate chapter of Darwall’s book includes an account of caring. Of the relation between caring and welfare Darwall states:

²⁰ The more specific, normative, version of what it is in fact rational to want for someone—the so-called “Aristotelian Thesis”—is spelled out in the final chapter of his book. What concerns me here though, is the so-called meta-ethical account of welfare and how caring fits into it.

In this chapter, we turn to the question of how to understand and identify the attitude that is featured in a rational care theory of welfare. This confronts us with the worry that it is impossible to define care or concern without already making use of the idea of a person's good or welfare, and therefore that we cannot define welfare in terms of rational care. (Ibid, pg. 50)

So Darwall takes himself to owe an account of caring that does not define it in terms of "well-wishing."

Perhaps surprisingly, Darwall's account of caring sidesteps the issue of circularity by claiming that he does not have to provide a definition of caring at all. On this Darwall says:

However, we need not define care (or, as I will call it in this chapter, sympathetic concern), if it is something like a psychological natural kind. Just as we can use a term like 'water' without a prior definition to refer to the natural stuff in the rivers and lakes for purposes of empirical theory, so likewise might we refer to care for purposes of a metaethical theory of welfare if it is a natural kind. (Ibid, pg. 50)

Darwall does not take himself to have to define care if it is a natural kind. Unfortunately, Darwall does not give criteria for natural kinds and does not really explain very well how the move to classifying caring as a psychological natural kind gets around the problem of circularity. I will try to explain Darwall's account.

Darwall contrasts sympathy with empathy (Ibid, pg 51-53). He says that whereas empathy consists in one person's feeling (or as least feeling something that resembles) what another person is feeling, sympathy is felt by a person from the perspective of another person. With empathy, one tries to "get into the shoes" of another, with sympathy, this is not the case. Care, for Darwall, is sympathetic concern. Darwall says of sympathy: "Sympathy, again, is a feeling or emotion that responds to some apparent obstacle to an individual's good and involves concern for him, and thus for his welfare, for his sake" (Ibid, pg. 67). Darwall offers repeated formulations just like this one. In Darwall's developed account of sympathy, there is still a reference to an individual's

good and her welfare. This might make the reader think that Darwall's construal of care makes his account of welfare circular. But to be charitable, Darwall must *not* intend the repeated formulation to be definitions; instead, they must be explanations of the psychological natural kind. Just what this amounts to, of course, remains to be discussed.

Before discussing it, I will mention one remaining aspect of Darwall's accounts of caring and welfare. Darwall takes caring (sympathetic concern) and well-wishing to have different objects. The object of caring is the individual. The object of well-wishing is states of affairs (Ibid, pg. 68-69). In Darwall's words:

According to philosophical orthodoxy, the standard object of desire, action and feeling is some proposition or possible state of affairs. If I want an ice cream, the real object of my desire is that I eat an ice cream. Or if I fear a tiger, then perhaps I fear that I might be eaten by a tiger. Moreover, it sometimes seems implicit in ethical writing that what it is to care about another person is simply or primarily to have a desire with a specific propositional content, namely, that the person fare well.

Even if desires and feelings have propositional objects, however, some also have "indirect objects" that are non-propositional. In particular, the form of desire involved in sympathetic concern does. Seeing the child on the verge of falling into the well, we don't simply desire that the disaster be averted. We desire this *for the child's sake*, that is, out of a sympathetic concern for *him*. (Ibid, pg. 67-68)

Darwall's proposal that care has, perhaps *inter alia*, a non-propositional object is intriguing. I think Darwall is right, but he has not sufficiently explained how his account does not face circularity. In what follows in this chapter, I will try to explain caring and well-wishing in a way that is friendly to both Darwall's and my accounts of welfare.

Let us return to Darwall's claims about natural kinds. A lot hangs on whether he needs to give a definition of "care." Whether Darwall can side-step the definition of care is the issue at hand.

5.3.3 Natural Kinds and Definitions

The natural kind term “water” picks out the natural kind water. Analogously, thinks Darwall, the psychological natural kind term “care” picks out the natural kind care. Presumably, so the argument must go, once care is established as a natural kind, it can be used in theorizing without providing a definition—for the denotation of the term “care” is set independently of definition. Here is my attempt at an illustration. We can imagine an elementary Earth Sciences class is learning the water cycle. This is oversimplified, but one could set up a chart with clouds, precipitation, soil, streams, rivers, and finally oceans linked in a cycle that is repeated until the end of the world. We can understand the functional relations between each of the components. We can understand that water *is* whatever has this functional role—independently of knowing whether water is H₂O.

On Darwall’s account by analogy, and this is somewhat speculative here, “care” is psychological natural kind term. We can understand the functional roles it plays in everyday lives even without a definition of “care.” We can understand its evolutionary history, its psychological and sociological import, and even its role in the account of well-being without providing an analysis of it. At least something like that seems to be going on in Darwall’s account of care and welfare. However, even if I am right up to this point, the issue is far from resolved. Certainly, Darwall ought to have said more about this.

Here is the quotation where it looks like Darwall might be giving a definition of “care,” but where I urged before that we not read him as doing such: “Sympathy, again, is a feeling or emotion that responds to some apparent obstacle to an individual’s good and involves concern for him, and thus for his welfare, for his sake” (Ibid, pg. 67). Instead of regarding this as a definition, let us call it a description of a functional state.

The description still involves the concept of welfare.²¹ It is not clear how we are to understand this description of the functional state and whether our understanding of caring is any better off than if the description were a definition. Let me therefore take a step back and examine for myself the relation between caring and well-wishing before explaining what I take to be Darwall's strategy of invoking natural kinds.

5.3.4 Sympathetic Concern and Well-Wishing

Why is the description of the functional state not problematic for us even though it makes reference to a person's good? Earlier we saw a couple of cases that show pretty convincingly that well-wishing is not the same thing as caring. The two cases were: (1) as case where I wish my brother well so that he will benefit me in turn, and (2) a case where the person is a Kantian and believes she has a moral obligation to at least sometimes benefit those in need and so desires to benefit someone in need solely from the motive of moral duty. In the first case we have instrumental well-wishing, while the second can be interpreted as either instrumental or intrinsic—depending on how we understand the motivation by a sense of duty. In neither of these cases can we say that the attitude is one of caring for the person. Caring and well-wishing are not identical. Moreover, even intrinsic well-wishing is not sufficient for caring.

There still remains the issue of whether caring is sufficient for well-wishing. If it is, then there still *could* be a circularity problem. It really depends on the nature of what I will call the "caring conditional:" if one cares for A then one well-wishes for A. Darwall, in describing sympathetic concern, repeatedly makes reference to an individual's good

²¹ Actually, Darwall does not quite say what he must mean in this quotation. It is not the apparent obstacles to an individual's good that provokes the feeling of sympathy. It must be the apparent obstacles *to what is judged to* be the individual's good that provokes the feeling of sympathy.

and welfare (Ibid, pg. 52, 67, 69, 70, 71, etc.). Now, the important thing to notice is that even if the “caring conditional” is true, it still could be that caring and well-wishing are not conceptually linked. If so, we could understand what care is without the concept of well-wishing even though every instance of caring is correlated with an instance of well-wishing. If this were the case, then there would not be a problem of circularity, even though the “caring conditional” is true. Though Darwall does not say it, maybe this is something like what he has in mind. It is a way of accepting the conditional and yet avoiding the circularity problem. The circularity problem is avoided, because even though each instance of caring is sufficient for a case of well-wishing, the concept of caring could be independently understood.

5.3.5 The Object of Care and Well-Wishing and “One Thought Too Many”

In this section I will try to explain how every instance of caring could have a corresponding well-wishing, while keeping caring conceptually distinct from well-wishing. If I understand Darwall’s position correctly, this is something that can help both his and my accounts of well-being. Two things, in particular, make me think that caring should not be analyzed as involving well-wishing: (1) the object of caring differs from the object of well-wishing, and (2) with well-wishing, there is one thought too many for it to be caring.

Darwall, as I mention above, thinks that the object of caring and the object of well-wishing differ. In his book he says:

When we care for someone we desire things for her *for her sake*. The object of care is the person *herself*, not some state or property involving her. In caring for her, we of course, want certain states and properties involving her to be realized. But when they derive from care, such desires also have an “indirect object” in addition to these direct objects. In caring for her, we want these things *for her* (Darwall, pg. 47, emphasis his).

What Darwall is getting at is that the object of caring and well-wishing differ in kind. In caring, the object is the person him or herself. In well-wishing, the object is some state of affairs. The state of affairs, presumably, must involve the agent, but the object is still a state of affairs. Darwall seems to rely on the suppressed premise that when the object of the mental state differs in kind, the mental state must differ in kind too. This seems like a fairly uncontroversial idea.

There is another, somewhat related, argument that should make us think that caring and well-wishing are not essentially related. Bernard Williams has offered a thought experiment that is helpful here. I think Williams' idea can be extended to caring and well-wishing. Williams, in his paper "Persons, Character and Morality," discusses the relation between personal connections and the demands of moral theories. In that paper he mentions the hypothetical case of a man who faces a dilemma. Both his wife and a stranger are drowning and he can only save one of them. He has a moral duty to save at least one. Some moral theories might recommend flipping a coin to determine which he should save. Yet other moral theories might recommend that he base his judgment on whatever facts he might know about the drowning victims. Perhaps, for example, one of them, if saved, will produce an enormous amount of happiness for others. But it is counterintuitive that one might be obligated to save someone other than one's wife. As a response to this, some defenders of some moral theories would try to show how the theories can make it at least permissible for the man to save his wife.

Williams sees it as a mistake to try to make the saving of the wife compatible with a theory. On this Williams says:

"But this construction [revising an account to make saving the wife compatible with it] provides the agent with one thought too many: it might have been hoped

by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife" (Williams, pg. 646).

The conflict is between love (or friendship) and moral duty. The person who thinks "It is my wife" and then jumps in to save her acts out regard for his wife. The person who thinks "It is my wife and in situations of this kind it is permissible to save one's wife" acts, at least in part, out of regard for moral considerations. So even if the defender of the moral theory can show that the theory is compatible with its being permissible to save one's wife, the defender still requires one thought too many. Requiring the extra thought, "In situations of this kind it is permissible to save one's wife," is counter-intuitive.

Williams writes about the feelings of love and friendship and how motivation from them is different from motivation from moral duty. I think that his argument can be modified for my purposes. I am not concerned with the feeling of love or friendship. Instead, I am concerned with care and what I call well-wishing. Well-wishing is analogous to motivation from duty, I say. When one gives advice based in well-wishing, one's state of mind is that A should have X and this is so because X contributes to A's well-being.²² Care, on the other hand, is analogous to love on Williams' account. The advice from someone who cares is that A should have X for A's sake. This might very well be what Darwall has in mind. In an example above, Darwall says that the state of mind of the well-wisher upon seeing that a child is on the edge of a well would be that

²² This is well-wishing for a particular person at a particular time. I suppose there is a more general kind of well-wishing—when one wishes that there is an increase in overall well-being in the world, but this general kind of well-wishing is not the kind of well-wishing that concerns me here. That general kind of well-wishing could not plausibly be thought to be connected in a strong way to caring.

disaster be averted.²³ But if I am right, then the state of mind of the well-wisher would be that the disaster be averted because this disaster would be against the well-being of the child.

Just as in Williams' case, the course of action recommended by the moral theory and the course of action from the motive of love is the same, so too can the advice from a well-wisher and the advice from a caring person be the same. However, just because the advice is the same, it does not follow that the two attitudes are conceptually connected. The "caring conditional" may very well be true, but this is a mere correlation.

There is a further complication on my view that Darwall does not face. As I discussed at length in Chapter 1, there is a narrow and a broad concept of well-being. On the narrow concept, well-being turns on the non-moral, non-aesthetic, self-regarding interests of the person. On the broad concept, well-being includes instead all categories of choiceworthiness in a life; so it could include moral values and whatever other values there might be. Now, to care for someone might involve desiring her broad well-being. An example might be of a parent who cares for her children and recommends, in virtue of this, that they be charitable. Now, this recommendation might very well result in a decrease in the narrow, prudential, well-being of the children. Restricting the advice to that which is agent-regarding and non-moral would work to eliminate this advice, but it is important to note that the advice from a caring person, without the "filter" provided in the earlier parts of this chapter, might not always be in the advisee's interests, narrowly construed, or rather, it might not further the advisee's narrow well-being.

²³ Of course, someone might have this general well-wishing state of mind. I suppose everyone might have it to some extent. However, this general well-wishing state of mind is not the salient one for my discussion, for it could not plausibly be confused with caring.

This point shows that my account of caring and well-wishing more closely analogous to Williams' account of love and duty than might have been thought. For in some cases, the course of action motivated by love will not be the same as that motivated by duty. And in my theory, in some cases, the advice motivated by care will not be the same as the advice motivated by well-wishing.

Given that I adopt the requirement that $A+$ care for A , it is important that I be clear on how my theory of well-being and Darwall's differs. First, Darwall's theory is a rational care view and not a fully-informed care view. It is not the case that one has to be fully informed to make rational, caring, judgments on Darwall's view; at least, he mentions no requirement for his theory of full information in his book and essays.

Second, I require that the advisor be the fully informed *counterpart* of the advisee. Darwall's theory has no such requirement. Any person, on Darwall's theory, suitably in a position to rationally care, can make judgment about someone's welfare. On my view, the fully informed counterpart has a privileged position for advising, because the advisor and advisee share a personality. Thirdly and lastly, Darwall heavily restricts what can count as "rational" in his rational care theory of welfare. He supports what he calls an "Aristotelian Thesis," according to which, one's life goes well when one appreciates and engages in activities worthy of merit (Darwall, pg 73-104). I support no such thesis.

5.3.5 Perverse Advice and Caring

Let me return to the worries with which I began this section. The worry is that if A hates himself, then $A+$, being in the nearest possible world in which he is fully informed, etc, will hate A and could advise him, say, to jump off a cliff. Desire-satisfaction theories, as a general class of theories, are going to be more permissive as to what they allow as making a person's life to go better or worse. But we should try to exclude so-

called perverse advice if possible. I think that the requirement of caring does the trick. Let me discuss a few examples.

If the advice to jump off a cliff springs from whim or fancy, then this advice can be eliminated. Care for someone precludes this sort of unconsidered advice. Care for someone requires that one take her plight seriously as opposed to treating it as an unimportant issue. So at least that kind of perverse advice is eliminated.

What, if after taking the matter seriously, the advisor who, we are stipulating, still cares for his advisee, nevertheless still advises him to jump off a cliff? Could this happen? In certain circumstances it might make sense to advise someone to jump off a cliff—such as if a much more horrible thing would happen as the only alternative. But the case I am envisioning (or trying to envision) is one in which there are perfectly fine courses of action as alternatives and yet the advisor's advice is to jump off the cliff. It is hard to make sense of this advice in such conditions and I am not sure it is intelligible if the advisor cares for the advisee.

Here is one possible line of response. The caring advisor, we can say, should take into account the desires of the advisee, especially those that are formed from good information. Now suppose, my opponent stipulates, that the advisee, himself, wants to jump off a cliff, even when he, the advisee, cares for himself. Such a case is quite absurd and it really is hard to understand the motives of such an imaginary person. The more the case is revised to get around the requirement of caring, the less intelligible the state of mind of the advisor and advisee gets. My ideal advisor account, fully developed, seems to handle objections from perverse desires reasonably well.

Indeed, my ideal advisor account handles perverse cases at least as well as other conceptions of well-being. It is easy to see how jumping off a cliff could contribute to one's well-being on mental state views. Say that the person can have pleasure only in jumping off a cliff and is simply miserable otherwise. Such a person may not care for himself and may take pleasure only in perverse things. In such a case, jumping off a cliff may be counted as contributing to the person's well-being on some popular mental state views. Mental state theories could perhaps be refined to exclude such cases. But I doubt it would be as easy as it is for the desire theory. So mental state theories likely fare worse than informed advice views on this issue.

Perfectionist theories do not fare worse, regarding perverse cases, than informed advice views, but they do not fare any better either. Perfectionist views treat a life that accords with and develops one's human nature as the best life. Now, like lemmings, it is possible for humans to have an impulse to jump off a cliff.²⁴ So too, in the right conditions, the perfectionist theory would have it that jumping off a cliff contributes to one's well-being. So, I do not see how perfectionist theories fare any better on this score.

5.4 Conclusion

In this chapter, I have discussed three areas where a generic ideal-advisor account, such as the one developed in chapters 3 and 4 of this dissertation, must be altered to fit the conditions of prudential well-being proposed in Chapter 1: Problems are posed by (1) desires that are non-self-regarding in that they do not seem to be relevant to an agent's life, (2) desires that seem to have moral and aesthetic content, and (3) desires that are self-regarding in the technical way discussed in 5.1 and 5.2, but that, intuitively, stem

²⁴ If the reader finds the example of lemmings implausible, as some have found it, I think other examples could be given.

from hatred or dislike of the agent whose well-being is at stake. I have dealt with each of the three sorts of problems in the sections above. I have argued that only a subclass of an ideal advisor's desires for her advisee are relevant to her advisee's well-being.

CHAPTER 6

MIXED THEORIES AND DESIRE-SATISFACTION CONCEPTIONS OF WELL-BEING

In this chapter, I will argue that my desire-satisfaction theory developed in this dissertation is superior to other accounts of prudential well-being. Let me explain why I *return* to critically evaluating conceptions of prudential well-being. In Chapter 2, I critically assess the prospects of desire-satisfaction accounts of well-being. In that chapter, I discuss three monistic conceptions of prudential well-being: mental state conceptions, desire-satisfaction conceptions, and objective human nature conceptions. The three conceptions just mentioned are monistic in that they each treat just one thing as constituting well-being. Additionally, in Chapter 2, I discuss so-called pluralistic conceptions which treat well-being as constituted by two or more things. I argue that desire-satisfaction accounts are superior to mental state, objective human nature, and pluralistic accounts of well-being. There are, however, other, more sophisticated, conceptions of well-being. In Chapter 2, I put off addressing some of the more sophisticated accounts of well-being. My idea in doing so was that an adequate comparison of these more sophisticated accounts of well-being with my preferred desire-satisfaction account requires that the reader know the details of the latter—the ideal advisor account—in a highly developed form. My goal in this chapter is to critically examine those more sophisticated accounts of well-being.

6.1 Mixed Theories: What They Are and What Motivates Them

A mixed theory is one that takes several things and treats each as a necessary condition for well-being. The important point at this stage is to understand that mixed theories *conjoin* conditions for well-being. An example of a relatively unsophisticated mixed theory would be one which takes happiness as a necessary condition for a person enjoying a degree of well-being while also taking the truth of the beliefs upon which the person's happiness is based to be necessary as well. An account that combines these two conditions would be a mixed account that combined a mental state component and, loosely speaking, a component having to do with states of the world. Generalizing, we can see that any mixed theory would have the following form: A person enjoys well-being (to degree X) if and only if, and because, the person has *both* characteristics (ϕ) *and* (ψ) (to degree X).

Mixed theories are different from pluralistic theories, such as those of Finnis and Scanlon.¹ Pluralistic theories, like mixed theories, combine different conditions in an account of well-being. However, pluralistic theories and mixed theories differ in that pluralisms treat each condition as *independently contributing to enhanced* well-being, whereas mixed theories treat each condition as *necessary* for well-being. A pluralistic theory would take the following form: A person enjoys well-being (to degree X) if and only if and because *either* the person has characteristic (ϕ) (to degree X) *or* the person has characteristic (ψ) (to degree X). An example of a simplistic pluralistic theory is one that treats friendship in a person's life as contributing to at least some well-being in his life and treats pleasure as contributing to at least some well-being. On this simplistic

¹ Please see Chapter 2 for a more extensive discussion of these pluralistic theories.

pluralistic theory, an agent's having some friendship or some pleasure would result in his having some well-being. I will take up mixed theories in this chapter.

What is appealing about a mixed theory is that one might hope that it would capture all of the attractive elements of each necessary condition while avoiding the defects of the conditions. Let me very briefly discuss the unsophisticated mixed theory just mentioned—the happiness-plus-truth theory. One of the problems with the happiness theory—as is seen from Chapter 2 of this dissertation—is that the happiness can be based on false belief. Experience Machine-type examples should have convinced the reader of that. One can try to overcome the counterintuitive aspect of the happiness theory by alloying it the requirement of true belief. Indeed, if this were the case, lots of counterintuitive aspects of the happiness theory would fall by the way-side.

6.2 Mixed Theories that Involve Desire-Satisfaction Accounts

Desire-satisfaction accounts, so we are lead to believe by their critics, have some serious problems. In what I have done thus far in this dissertation, I hope to have convinced the reader that many of the issues thought to afflict desire-satisfaction accounts of well-being do not really do so, at least they do not afflict the best desire-satisfaction theory. There are, however, a few potential problems I have not discussed. One might think that desire-satisfaction accounts can be improved by adding requirements. The resulting theory would be a mixed theory on my taxonomy.

Here is one potentially problematic case that might make one think that desire-satisfaction accounts of well-being could be improved by additional conditions. There seem to be certain kinds of desires that could be in a person's so-called relevant desire set but yet their satisfaction should not, intuitively, enhance her well-being. An example

found in the literature is the desire to turn radios on and off.² This desire may meet the condition of being self-regarding as explained in Chapter 4, because what the person desires, may be that *she* turn radios on and off continually. There is no reason for thinking that ideal advisor accounts get around this problem, at least at first glance. Perhaps, however, one could deal with this problem by developing a mixed theory of well-being: for example, a theory that treats well-being as constituted by the satisfaction only of desires that have, as their objects, things which conform in some relevant way with our human nature.

Let me briefly explore another potentially problematic case. There is thought to be something counter-intuitive about allowing mere desire satisfactions to constitute well-being, without requiring any awareness of the desire satisfaction on the part of the agent. Recall that a desire in someone's so-called relevant desire set can be satisfied and she not know it. This supposed weakness of desire-satisfaction theories, as they are traditionally understood, can perhaps be countered by adding a necessary condition found in some other conception of well-being, such as a mental state conception or a perfectionist conception —thus producing a mixed theory.

I will argue in what follows that neither of these supposed weaknesses in desire-satisfaction accounts is as significant as critics have thought and that adding an extra necessary condition certainly would not improve upon my ideal advisor theory. For simplicity of expression, let us say that a mixed theory that involves desire-satisfaction

² This example is from Warren Quinn's "Putting Rationality in its Place" as found in his book of collected essays, *Morality and Action* (pg. 236). In that essay, Quinn seems to be concerned with arguing that "subjectivism about moral value" is false (pg. 228-229). He might not have formulated his example specifically to argue against desire-satisfaction accounts of prudential well-being, but I will nevertheless use his example as though it were so formulated.

has the general form: A person enjoys well-being (to a certain degree) if and only if and because both (i) a desire in the person's relevant desire set is satisfied and (ii) some other non-desire-satisfaction-related condition is met. I will discuss in some detail several specifications of (ii).

6.2.1 Desire/Objective Mixed Theories

Let me first deal with examples that have to do with desires that are in the relevant desire set on at least some desire-satisfaction accounts and yet, intuitively, the satisfaction of them does not seem to enhance one's well-being. There is the example of a person who desires that she turn radios on and off. Another example that appears in a number of works is of a person who wishes that he count blades of grass.

One way of trying to fix up this problem is to set restrictions on what desires can count as relevant to one's well-being. One could combine a desire-satisfaction conception of well-being with a perfectionist conception of well-being—resulting in a mixed theory. The resulting theory would be a mixed theory, according to which one's well-being is constituted by the satisfaction of desires one has, the satisfaction of which, say, would develop or exercise one's human nature. The theory under discussion could be stated as follows: A person enjoys well-being (to degree X) if and only if (i) a desire in the person's relevant desire set is satisfied and (ii) the satisfaction of this desire involves the exercising or development (to degree X) of the person's human nature. Call this the "Mixed Desire-Satisfaction/Human Nature" theory or "MDH" for short.

According to this mixed view, one might argue, the desire of a person that he count blades of grass would not be relevant to his well-being. Now, it is not clear whether the proponent of the mixed view would be right about this. Part of the problem is that scholars have not provided much detail as to what sorts of activities (putatively) would

develop or exercise human nature. Nevertheless, as I discuss in Chapter 2 of this dissertation Hurka claims that an activity that develops or exercises one's physical nature or one's practical or theoretical reasoning enhances one's good—for my purposes, I will take what he says about the human good as though it were offered as an account of prudential well-being. It is not clear whether Hurka's account of human nature could help with the problematic case mentioned above. Perhaps counting does exercise one's reasoning. In any case, let us assume that an appeal to human nature does solve the problem for pure desire-satisfaction accounts of well-being.

I do not know of any philosopher who has whole-heartedly endorsed the above-mentioned mixed view; however, several have flirted with it. For example, T.M. Scanlon, in his *What We Owe to Each Other*, entertains a view that has some similarities to the mixed view I have been discussing. Scanlon finds the informed desire view attractive but subject to serious criticism. He then proposes a “rational desire” view of well-being (even this, ultimately, he does not endorse). Scanlon explains the sense of “rational desire” that he has in mind as follows: “by a person's rational aims we might mean aims that he or she actually has, insofar as these are rational (that is to say, insofar as the nature of these aims does not provide good reason to revise or abandon them)” (Scanlon, pg. 119). The idea that Scanlon has is to bring into the picture some ideal of worthwhile aims, apart from their being merely the object of desire. On this Scanlon says:

The requirement that an aim be rational incorporates this critical element by allowing for the possibility of substantive criticism of aims. This requirement also accommodates the fact that from an individual's own point of view what makes an aim worth adopting and pursuing is, first and foremost, not merely its being chosen or desired but the considerations that (in his or her view) make it worthwhile or valuable” (Ibid, 119).

The proposal that I understand Scanlon to be considering is that we accept the desire view in one of its forms—it seems to be the actual desire view—but that we place constraints on which desires can count as relevant to one's well-being. These constraints are based in a conception of independent value.

Perhaps surprisingly, Scanlon does not say much about what these substantive independent values are supposed to be. Later, Scanlon discusses how the rational desire view can deal with some criticisms of so-called objective accounts (Ibid, pg. 120).³ Scanlon distinguishes between subjective and objective accounts of well-being, following James Griffin in his *Well-Being*. Griffin includes basic needs accounts and objective list accounts as types of objective accounts. I assume that he would also include perfectionist accounts in that category.⁴ We cannot be sure what Scanlon has in mind, but my thinking is that since he makes clear reference to objective theories, on Griffin's construal of what objective theories are, Scanlon must have in mind a mixed theory that is part desire theory and part objective theory.

I will spend the next two paragraphs sketching more of the details of Scanlon's account of well-being so that the reader has a better idea of how the mixed conception fits into his overall picture. Then I will return to the mixed aspect that I think has some initial attraction. Scanlon's ultimate account of well-being is pluralistic, one part of which is mixed. He takes rational aim satisfaction to be one potential constituent of well-

³ Although I approach the mixed theories from the side of the desire-satisfaction views, there are other relevant perspectives—such as from the standpoint of the *other* theory. Objective accounts are often thought to be too inflexible and to specify that one way of life is best for everyone. The mixed theory under discussion would answer this criticism of objective theories by adding flexibility and choice as aspects of the account of well-being.

⁴ Griffin, like Sumner, adopts a two-part taxonomy: the subjective and the objective. This is different from my preferred three-part taxonomy. Had Griffin considered perfectionist accounts of well-being, he most certainly would have counted them as objective accounts.

being. Also, he takes certain experiential states, for example, enjoyment, to be a second potential constituent of well-being. “Thirdly” he says, “many goods that contribute to a person’s well-being depend on the person’s aims but go beyond the good of success in achieving those aims. They include such things as friendship, other valuable personal relations, and the achievement of various forms of excellences, such as in art or science” (Ibid, pg. 125). Scanlon’s third condition is a confusing and he does not try to explain it. I am not sure if he means that there are also instrumental goods that come about from achieving one’s rational aims or something entirely different.

Scanlon’s account falls victim to the same problem that plagues other pluralisms—that of commensurability between the various parts of the account.⁵ Scanlon’s pluralism is a lot like the other pluralisms I discuss in Chapter 2 of this dissertation. Each item on the list has some initial plausibility but also adds further complications. Furthermore, we are not given a way, even in principle, of making well-being assessments for any given individual. There is no principle of commensuration that would allow for there to be, even in principle, a total overall amount of well-being one has that results from taking the sum of the amounts of well-being one enjoys in each of the categories of the pluralistic theory. So Scanlon’s overall view can be rejected for the reasons I provided in arguing against pluralistic accounts in Chapter 2.

But let us return to what I think is an intriguing idea—leaving aside Scanlon’s pluralism. Let us take the rational desire view as a mixed view, which combines a desire-satisfaction aspect with either a perfectionist aspect or a pluralistic aspect. To revert to

⁵ In Scanlon’s defense, he admits he is not offering a unified “theory” of well-being (Ibid, pg. 125). He thinks that we are “unlikely” to find a unified theory of the sort I am after in this dissertation. I do not see any reason to be so pessimistic.

the language found above, the theory could be stated as follows: A person enjoys well-being (to a certain degree) if and only if and because *both* (i) a desire in the person's relevant desire set is satisfied *and* (ii) the object of this desire is a rational aim for him. Call this the "Mixed Desire-Satisfaction Rational Aim" theory or "MDR" for short. As I say above, Scanlon does not say much about the criteria that he thinks a rational aim would have to meet, but given the context, it appears that he has in mind that the an aim would be rational if it corresponded in the right way to an item on an objective list of some sort. Let us treat MDR as holding that the object of a desire is a rational aim just in case it is on the list. This mixed view might solve some of the problems of typical pure desire-satisfaction views.

Now, it seems that MDH and MDR deal with problematic cases for pure desire theories in an intuitively attractive way—at least in a limited range of cases. Moreover, the addition of the desire component to a pure perfectionist account could make the resulting theory more flexible than the pure account, and its addition to a pure objective list account could solve the commensurability problem that afflicts pure objective list views. Perhaps, for example, strength of desire could be used to make commensurable the items on the list.

While there is some initial plausibility to such mixed views, they ultimately must be rejected. One of the mixed theories, MDH, combines a desire-satisfaction element with a perfectionist restriction on the relevant desire-set. Such a theory faces problems. For starters, the addition of the desire-based aspect still does not help us get a fix on what our human nature is. That enormous problem still remains. The most plausible account of human nature seems to be the evolutionary account, but, as I argued before, it does not

yield an adequate perfectionist account of well-being, even if we grant that it is an adequate account of human nature. Please see chapters 1 and 2 for details on this point. Even where the relevant desire set is restricted to desires that are appropriately related to the exercise and development of human nature, insurmountable problems remain.

The mixed theory that combines desire-satisfaction and an objective list, MDR, is a little bit better than the desire/perfectionist account, but still, I suspect, ultimately untenable. The primary problem with the pluralisms we have seen so far is the commensurability issue. Perhaps strength of desire can solve that problem, but I think that there is a dilemma about commensurability that the proponent of MDR must face. Either (1) we do not solve the problem of incommensurability or (2) we do. If we do not, then the problems we may solve by adding the objective list aspect to the desire aspect are not important enough to increase the overall plausibility of the pure desire-satisfaction theory. This is because the core problem for pluralistic theories is that they have the problem of incommensurability—as I discuss in more detail in Chapter 2 of my dissertation. If we do solve the incommensurability problem, then new problems arise. The problems, if the theory falls on the first horn of the dilemma are straightforward and so I will not go into a discussion of them. I will explore the second horn of the dilemma.

My suspicion is that we could use strength of desire to solve the problem of commensurability. The rough idea is that each individual is faced with a list of goods, and it is the strength of his desires that determines the amount by which his well-being increases when he obtains different goods. This seems to solve the problem at least “in principle” which seems sufficient for my purposes. I suspect though, that if we use

strength of desire to solve the problem of incommensurability, then a new problem will show up. Let me explain.

Thus far in my dissertation, I have allowed those who would support an objective list theory to put anything on the list that he or she likes. This is important because I have not required an account of where the list comes from or how it is generated. My suspicion, however, is that if we use strength of desire to solve the problem of incommensurability, and do not provide an account of how the list is made or where it comes from, then the desire component will “take over” the theory.

I will explain what I have in mind by an example. Let us say that there are two items on the list: aesthetic appreciation and friendship. Furthermore, let us say that there is some agent with three desires: a desire for aesthetic appreciation, another desire for friendship and a third desire for pleasure. The strongest desire is for aesthetic appreciation, next is the desire for pleasure, and the weakest of all is the desire for friendship, let us suppose. According to the kind of theory we are considering, the satisfaction of the desire for aesthetic appreciation would lead to a great increase in well-being; the satisfaction of the desire for friendship would lead to a relatively small increase. When asked why there is this difference, the proponent of MDR would say that strength of desire makes commensurable the various items on the list and determines the degree to which they contribute to well-being. But if the desire-component of MDR is as significant as this, we must wonder why the desire for pleasure does not make pleasure relevant to the agent’s well-being. A proponent of MDR will no doubt take it that this is explained by the fact that pleasure is not on the list, while friendship and aesthetic appreciation are. This explanation is unsatisfactory, however, unless some account can

be given of why pleasure is not on the list in the first place.⁶ If the proponent of MDR had a story to tell about the generation of the list, matters would be different, but so far, such a story has not been provided by those who have proposed list theories. If no satisfactory explanation can be given, then it is not clear why a desire, by itself, does not make its object relevant to well-being. The existence and strength of desire, after all, already plays a primary role in theory. Absent an explanation of what determines the content of the list, it is far from clear whether the list makes an important contribution to the plausibility of the theory.

Recall at this point that we have been discussing the example of a person who has a desire to count blades of grass, and the example of a person who desires to turn radios on and off, on the assumption that the examples result in counter-intuitive implications for at least some pure desire-satisfaction accounts. Given this assumption we have been exploring ways to amend such accounts. The important question to ask at this point is whether the examples are as problematic as they originally seemed. I think they are not. To some extent, I think that my ideal advisor theory deals with this sort of example in an intuitive way.

Firstly, we must be sure that, when considering such examples, we keep in mind that the desires in question must be someone's desire that *he* or *she* does something. If the desire is of the general form, *A* desires that *A* do *x*, then we can think of the example as being concerned with a *personal project* of sorts. Admittedly, it is hard to see how someone could have a personal project that he count blades of grass, but then there are

⁶ Whether or not this demand for explanation should be pushed upon the pluralist is another matter. I did not demand an explanation for the list in the second chapter of this dissertation. At that point, I ignored the question as to what makes the list. It still strikes me that the demand for explanation is especially forceful in the context of the mixed theory under consideration.

many people who cannot “make sense” of many other people’s projects. For example, many people cannot understand why others collect things, such as stamps or coins. If we were to require that A’s personal projects make sense to others in order for them to be relevant to A’s well-being, then perhaps many things that the reader holds dear would be excluded from counting as relevant to his or her well-being. That seems, I hope, counter-intuitive.

Secondly, I think that examples such as the desire of a person that she turn radios on and off might not meet the conditions my theory places on the relevant set because turning radios on and off is not the sort of thing that an ideal advisor would advise herself to do, were she to care for her less informed self. I will not explore this reply, but think that it is possible that the requirement that the ideal advisor care for her actual self could rule out an advisor’s advising her advisee to turn radios on and off or to count blades of grass.

In sum, I do not think that either MDH or MDH solve the original problem for pure desire theories. Furthermore, I do not think that the original problem is as serious as it may initially have seemed—for I do not think that the examples provided lead to counterintuitive results. On careful reflection, it seems that my preferred desire-satisfaction account may be able to handle the problem.

6.2.2 Desire/Mental State Mixed Theories

As was mentioned above, a feature of the desire-satisfaction account of well-being is that a desire can be satisfied and yet the agent never knows about it. This is thought to result in counter-intuitive results for desire-satisfaction accounts. The general concern is that one’s well-being, according to such a theory, might be too disconnected with the person’s lived experiences.

One way of trying to solve this problem is to add a belief component to the original theory. I do not know of any philosopher who has proposed this approach, but there have been a few who have considered it, such as Griffin in his *Well-Being*, and others.

I will briefly discuss two difficulties with this new mixed view and then I will turn to a more serious objection to the idea that we need to add a belief component to pure desire-satisfaction theories. One of the difficulties is that it is not clear which belief component should be added. Recall that the general form of a mixed theory with a desire component is as follows: A person enjoys well-being (to a certain degree) if and only if and because both (i) a desire in the person's relevant desire set is satisfied and (ii) some other non-desire-satisfaction-related condition is met. The issue at hand is to identify which belief component should fill in for clause (ii). There seem to be at least two candidates; (a) require the person to believe that the desire in the relevant desire set that is mentioned in (i) is satisfied or (b) require the person to believe that p where the desire that p is the desire in the relevant desire set that is mentioned in clause (i). Call the mixed theory that replaces clause (ii) with (a) the "Mixed Desire-Satisfaction/Belief Satisfied" theory or "MDBS" for short. Call the theory that replaces clause (ii) with (b) the "Mixed Desire-Satisfaction/Belief Proposition" theory or "MDBP" for short. I think that there are serious problems with both MDBS and MDBP. These problems are significant, even in theories that provide counterfactual accounts of the relevant desire set. Let me explain.

The best accounts of the desires that are relevant to well-being are counterfactual—as I have argued extensively in chapters 3 through 5 of this dissertation. The move from actual to counterfactual desires has a striking result for the mixed theory. On counterfactual accounts of the relevant desire set, the agent whose well-being is under

consideration will not always actually have a desire that belongs in the relevant desire set. When the agent does not actually have such a desire, it is hard to see how a belief requirement could possibly make the theory give more plausible results. With MDBS, it is clear that there could be cases where many agents do not know what their desires would be under the counterfactual conditions specified by the theory. Also, with MDBP, there will be cases where the agent would believe that p and yet never know that the proposition p is the object of a desire which is in the relevant desire set. What I have said here is too speculative to undermine all possible desire/mental state accounts. However, the considerations mentioned should definitely make us think that any desire/mental state theory will face complex difficulties.

There is a more devastating objection to desire/mental state theories. Desire/mental state theories do not actually solve any problem with the desire theory. The first challenge to any proponent of a mixed theory is to show how it solves a problem with a pure theory. Critics of desire-satisfaction theories say that such theories are faulty because of the possible gap between actual satisfactions and the mental life of the agent. For example, Sumner suggests that “[a] theory of welfare can be descriptively adequate only if it incorporates some form of experience requirement; this was the important insight in classical hedonism” (Sumner, pg. 128). Sumner seems here to say that a necessary condition of an adequate account of welfare or well-being is that one “experiences” the grounds of one’s well-being. I suggest that a better way of characterizing the new requirement is as a “belief” requirement. Nevertheless, Sumner clearly seems to think that there is a serious problem with pure desire-satisfaction theories. I disagree with Sumner and other critics wholeheartedly. There is no need to

add a belief requirement to pure desire-satisfaction theories, I say, because there is not a problem of the sort critics like Sumner envision to begin with. Desire-satisfaction theories are far more resourceful than their critics have supposed.

Consider a slightly modified version of an example that comes from Parfit, the example of a person who wants his children to do well. We can imagine that they do well and yet the parent never comes to know this. In fact, we can imagine that the parent is far removed from the lives of his children and does not know anything about them. Now, presumably, the desire that his children do well is relevant to his well-being. At least, the position I argue for in chapters 4 and 5 would make the desire relevant. The objection we are considering is that the parent's lived experience is far separated from his well-being on pure desire-satisfaction accounts.

In thinking about this case, we bring to the table certain presuppositions that might make the case look problematical for desire theories. Once these presuppositions are fully laid before us for consideration, however, the case does not turn out to be problematic. We can imagine ourselves, or others we care for, in a very similar situation to the parent. In such cases, we might think that our life would go better if we knew about the desire-satisfaction. This much I think is right. However, when we think of such cases, we also have in the background the thought that the parent wants *to know* what happens to his children. It is sensible to assume this in the case, i.e., to fill in the details in this particular way. Now, however, we have put into the case another desire—the parent's desire *to know* whether his children do well. Of course, this is an ordinary desire and it can be satisfied or not. In the hypothetical case this additional desire is not satisfied. But now we have something to point to when we claim that the parent's life

does not go as well if he never knows that his children do well as it would if he did know this. The additional desire is nothing special and requires no addition to simple desire-satisfaction theories. Our intuitions about the hypothetical case are right, I say, for we are assuming that the parent desires both that his children do well and that he knows that his children to well. Once we make this assumption explicit, we can explain the case perfectly.⁷ There is no objection here to my theory.

Countless cases can be handled in this way. This is because most of us think that knowledge is highly desirable. However, there will be hypothetical cases that favor the pure desire-satisfaction account and make the mixed theories, such as MDBP and MDBS, look mistaken. Although knowledge of the conditions of one's life and the surrounding environment is often seen as extremely desirable, there are cases where it might not be. For example, some people wish not to know the sex of their fetus. I am not sure why this is, but let us consider cases in which parents have an intrinsic desire not to know. Among those who want not to know the sex of their fetus (at least until it is born), some will have a desire that the fetus be, say, a female. On the desire/belief views I have been examining, the "expecting" adult's well-being will be increased if the fetus is indeed female and the adult has the relevant belief. This seems very counterintuitive indeed. For it is far more plausible to think that the adult is better off if the fetus is female and even better off not knowing the sex at all, given that he or she has a desire not to know.

⁷ An interesting thought experiment is to think of the hypothetical case with two different modifications: (1) the case is the same but the parent *does not want to know* what happens to his children or (2) the case is the same as the original, but the parent *wants not to know* what happens to his children. Colloquially, when people say "don't want . . .," they often just mean "want not . . ." Here I am using the terms a bit more rigorously. There is an absence of a desire in (1), and a desire is present in (2). These are hard cases and we will not have as strong an intuition as we do with the original, unaltered, case. With the first change, desire satisfaction theories treat the knowledge as irrelevant to one's well-being. With the second change, desire satisfaction theories treat knowledge as *decreasing* one's well-being, ignorance as *increasing* it. These are hard cases and both results seem acceptable.

The hypothetical case where the parent has a desire to know about his children's well-being should make us think that the desire/belief view is an unnecessary complication—that the pure desire view does a good job accounting for our intuitions in the case. The case of the “expecting parent” should make us think that the mixed theory is simply mistaken where a desire theory does a fine job accounting for our intuitions. So, although philosophers have thought that there is a serious problem with desire-satisfaction theories of well-being because they allow there to be a gap on the theory between one's well-being and one's conscious life, this turns out not to be the case.

6.3 Mixed Theories that Include Mental States

While there are few proponents of mixed theories that include desire satisfaction as an essential component, there are more proponents of mixed theories that include happiness as an essential condition. Several philosophers *entertain* the idea of mixing the desire-satisfaction account with some other theory, but few actually *endorse* such theories. Mixed theories that include happiness are more commonly endorsed. Perhaps the best-developed account of a mixed theory that includes happiness as a necessary condition, indeed, the best mixed theory offered of any sort, is that of Wayne Sumner in his book *Welfare, Happiness and Ethics*. It is to his account of well-being that I will now turn.

6.3.1 Sumner's Mixed Happiness Theory

After a survey of hedonism, desire-satisfaction theories and a few objective theories Sumner finds significant faults with each. He then explores various mixed theories including a desire satisfaction theory with an added experience requirement, which is somewhat like MDBP and MDBS. Finding fault with mixed theories that include a

desire-satisfaction component, Sumner then goes on to develop and endorse a mixed theory that includes happiness as a necessary condition.⁸

Sumner gives an admirably straightforward summary of his position in the following quotation:

The theory I shall defend does not simply identify well-being with happiness; additionally, it requires that a subject's endorsement of the conditions of her life, or her experience of them as satisfying or fulfilling, be authentic, the conditions for authenticity, in turn, are twofold: information and autonomy. Welfare therefore consists in authentic happiness. This theory is subjective, since it makes a subject's welfare depend on her attitudes, and since the function of the authenticity requirement is to ensure that these attitudes are genuinely hers. It satisfies the experience requirement, since a subject's happiness is a matter of her experience of the conditions of her life. However, it is not a mental state theory since authenticity is a relation between the subject and the world. The happiness theory thus mediates between hedonism and the desire theory, exploiting the strength of each while avoiding their weaknesses. (Sumner, pg. 139)

In the first sentence of the quotation, Sumner is alluding to his thesis that a particular kind of happiness is important for well-being—he describes this kind of happiness variously as life satisfaction or happiness about how one's life is going.⁹ Note that “Satisfaction” refers to a mental state on Sumner's account, but on the desire-satisfaction account, “satisfaction” refers to whether the object of the desire comes about. But even the restriction to “life happiness” is not sufficient in Sumner's view. Additionally, this “life satisfaction” must be authentic. I will spend much more time below talking about what Sumner takes to be two elements of authenticity: information and autonomy.

⁸ I am confident that I have already addressed Sumner's criticisms of the desire-satisfaction conception of well-being and I have already discounted mixed theories that combine desire-satisfaction with an experience requirement—or at least mixed theories which are very similar. In fact, I have a hard time seeing how Sumner would flesh out, in a plausible way, the mixed theory that has a desire component and an experience component, except by fleshing out the details in a way that is identical to either MDBP or MDBS

⁹ There are other types of happiness that are either irrelevant or at least less relevant to one's welfare, such as happiness about others and so on. By focusing on life happiness, Sumner is trying to avoid the problem of kinds of happiness that are irrelevant to one's welfare. There is an analogous problem with afflicts the desire-satisfaction theory which is discussed in chapters three through five in this dissertation.

Before leaving my discussion of the quotation above, however, it is worth mentioning that perhaps Sumner would have done better to say that his theory is not *purely* a mental state theory instead of simply saying that “it is not a mental state theory.” His final account of well-being does have a mental state as a necessary condition, though the mental state must relate to the non-mental world in the right kind of way. Thus, his account is best viewed as a mixed theory.

Although there are interesting issues about the happiness component of Sumner’s theory, such as the issue whether life satisfaction is best thought of as a kind of happiness as opposed to some other mental state, I will focus instead on what is to supplement the relevant mental state.¹⁰ In order for happiness to be authentic, we are told, it must be informed and autonomous.

I will now turn to the two conditions of authenticity on Sumner’s theory. One of the central problems with pure mental state theories, such as happiness theories, is that they do not have any connection to the real world.¹¹ Because someone can be completely mistaken in her beliefs about her circumstances, she can have the relevant mental state, and yet, intuitively, still have a very low level of well-being. If a restriction is placed on the mental state which, in effect, ties it to the world, then one of the central worries can be addressed. Sumner calls such a restriction an “information requirement.”

¹⁰ I do have some doubts as to whether the mental state picked out by Sumner is indeed a kind of happiness, but I certainly agree that life satisfaction is a mental state. The central issue which concerns me is not whether the mental state under discussion is happiness, but whether it fits well into a theory of well-being. So my focus is on what is mixed in with the mental state, regardless of whether the mental state is indeed a kind of happiness.

¹¹ Of course, one can maintain that happiness is not purely a mental state (some say Aristotle thought this). But this is not at issue here because a person who develops a theory of well-being that uses such a conception of happiness will be endorsing a mental state theory alloyed with something else.

Sumner first looks at three candidate requirements, rejecting the first two and accepting the third. Of the first he says: “The strongest candidate would be a *truth or reality requirement*, which would stipulate that happiness counts as well-being only when it is based on a view of the conditions of our lives which is free from factual error” (Sumner, pg. 158, italics his). Another way of putting this condition is that for happiness to count towards one’s well-being, it must be based on true belief. Sumner considers the reality requirement to be unreasonably puritanical (Ibid, pg. 158). He says:

When we reassess our lives in retrospect, and from a superior epistemic vantage point, *there is no right answer to the question of what our reaction should be*—that is surely up to us. Because a reality requirement stipulates a right answer—any happiness based on illusion can make no intrinsic contribution to our well-being—it must be rejected as presumptuously dogmatic. (Ibid, pg. 158-159, italics his)

Sumner thinks that the reality requirement is too strong in that happiness that is not based on “reality” is automatically discounted as not relevant to well-being. The reality requirement undermines individual sovereignty, he says, because some people would still think that their lives were going well even under the belief that what they thought in the past was incorrect. For example, one could look back at time spent in a Nozick-style experience machine as increasing one’s well-being (Ibid, pg. 161, especially footnote 25).

The second information requirement Sumner considers is a justification requirement. The justification requirement is weaker than the reality requirement in that it requires only that one be justified in the beliefs upon which one’s happiness is based, which falls short of requiring truth, because one can be justified in believing something without it’s being the case. Sumner finds the justification requirement slightly more plausible than the reality requirement, but he still finds the justification requirement too “arrogant” (Ibid, pg. 159). On this Sumner says: “Once again it presumes to dictate to individuals how much their deviation from an ideal epistemic standpoint should matter to

them. But that is for them to decide” (Ibid, pg. 159). So, Sumner rejects the justification requirement for the same reason that he rejects the reality requirement, that it does not leave well-being up to the agents’ decision in the right way.

Having rejected the reality requirement and the justification requirement, Sumner proposes his favored version of the information requirement. He suggests that the line of argument go in a “somewhat different direction” (Ibid, pg. 159). He says:

After all, what we are seeking is an adequate subjective theory of welfare, one on which the subjects’ point of view on her life is authoritative for determining when that life is going well *for her*. By connecting welfare with happiness we have interpreted that point of view as an endorsement or affirmation of the conditions of her life. When that endorsement is based on a clear view of those conditions, we have no grounds for questioning or challenging its authority: in this respect, the individual is sovereign over her well-being. But when it is based, wholly or partly, on a misreading of those conditions then its authority is open to question, since it is unclear whether or not she is endorsing her life *as it really is*. (Ibid, pg. 160, italics his)

The details of Sumner’s proposal are a little unclear in this quotation. He seems to add the requirement that one would “endorse the conditions of her life” were she to be informed of the matter. Briefly, Sumner seems to favor the following theory: A person enjoys well-being (to a certain degree) if and only if and because both (i) the person is enjoying a mental state of life satisfaction and (ii) the person would approve of the conditions giving rise to, her enjoyment of life satisfaction were she to be informed of the matter. I have already spoken of Sumner’s view on what “life satisfaction” consists in. Let us leave aside for the moment the question of what endorsement could be on Sumner’s proposal. I will briefly explore the information requirement Sumner seems to favor.

He thinks that information is relevant to A’s well-being “whenever it would make a difference to a subject’s affective response to her life, given her priorities” (Ibid, pg.

160). We can imagine someone who is quite satisfied with his life and who has the relevant mental state—life satisfaction—in abundance. But let us suppose that the person is deeply misinformed as to the real circumstances of his life. On Sumner’s preferred version of the information requirement, the person’s life happiness counts as well-being only when correcting the misinformation would *not* make the person somehow disqualify the happiness from being relevant to his or her well-being. Of course “somehow disqualify” is quite vague and I will explore a few ways of clarifying the terms.

Sumner’s information requirement can be expressed more clearly with a counterfactual. If an agent were to endorse the conditions of his life were he to be informed in the right way, then his relevantly related life satisfaction constitutes at least some degree of well-being.

Even though Sumner does not say he endorses a full information account of his information condition, I do not see how Sumner could *avoid* adopting a full information view. He suggests that “The relevance of information for a person’s well-being is a personal matter to be decided by personal priorities; there is here no authoritative public standard” (Ibid, pg. 161). Sumner *seems* to be saying that each individual gets to choose whether information is relevant to her choice or not. If that is what Sumner has in mind, this remark by Sumner should be disregarded. For an agent may have different answers regarding which bits of information are relevant given different informational conditions. There is no reason for choosing one over the other, if we leave the matter up to personal preference. We know he needs to have all of the information *relevant* to the life satisfaction in question. Absent an account of which epistemic viewpoint is superior in regards to an individual’s endorsement of her circumstances, the agent in the

counterfactual condition might have to be fully informed.¹² After all, we could imagine a situation in which the agent has some bit of misinformation corrected, but some other bit might just make the difference. This is a perplexing issue for the interpretation of Sumner's position, but as I argue above in chapters 3 – 5, the concept of an agent's being fully informed is not seriously troubling and Sumner could simply make use of the account I develop in those chapters.

We seems to have a fairly clear idea of what, on Sumner's view, life satisfaction and the information condition amount to, but we still have not explored the details of the requirement that *one endorse the conditions of her life*. I can think of three different accounts of what Sumner may have in mind: an agent, A, would endorse the conditions of her life when given the relevant information if either (a) A would believe that her life satisfaction is relevant to her well-being if she were relevantly informed or, (b) A would remain happy if she were relevantly informed or, (c) A would desire when relevantly informed that the things that actually ground her life satisfaction give rise to her sense of life satisfaction. In a number of passages, Sumner seems to support (a), but I will argue that this is not the best interpretation of his position because (a) leads to a vicious kind of circularity. Sumner seems not to think that (b) is the proper way to understand what endorsing the conditions of one's life amounts to, but I think that (b) is, nevertheless, worthy of being explored as an option. I think that (c) is the most charitable way of fleshing out what Sumner has in mind. I should make it clear to the reader, however, that

¹² The ultimate solution to what information is relevant and what is irrelevant cannot invoke yet another counterfactual, e.g., information is relevant if it would make a difference in judgments about what information should make a difference, for then the solution just passes the problem onto another counterfactual situation. Perhaps a better way to put this is to say that the way of distinguishing whether information is relevant or not cannot ultimately be solved by invoking yet another counterfactual. At least, it seems that this line of argument would eventually result in a full information account.

I think that Sumner's position, even when fleshed out in the most charitable way possible, is still seriously and fatally flawed.

Let us explore (a). On this option, an agent considers, with relevant information, whether some instance of life satisfaction is partly constitutive of his or her welfare. In many passages, Sumner seems to have such a construal in mind. For example, Sumner says that there are two options one can take once undeceived about the faithfulness of one's partner. Here they are:

One is that she re-evaluates *how well her life was going* (not how happy she was) during the period of deception: 'I thought everything was going *so well*, but now I can see that it was all a farce.' In that case, the discounting rate she now imposes on her earlier assessment of her well-being determines how relevant the information was. The other possibility is that she does not care: 'C'est la vie; at least he was charming and we had a lot of fun.' Here the information turns out to have zero relevance, since that is the status she confers on it." (Ibid, pg. 160-161, italics added)

It *seems* in this passage, that Sumner has in mind that the object of endorsement is one's life and endorsement, on this construal, involves a judgment by one of his or her level of well-being. The above is not the only passage that suggests (a). Sumner says of each individual that: "Their self assessments are therefore determinative of their well-being unless they can be shown to be inauthentic, i.e. not truly theirs. The requirements that these assessments be informed and autonomous spell out the conditions of authenticity" (Ibid, pg. 171). There are numerous other passages that provide at least some support for reading (a).¹³ I will argue, however, that (a) cannot be what Sumner has in mind, even if several passages seem to support such an interpretation.

¹³ Perhaps most notable is the passage where Sumner says: "After all, what we are seeking is an adequate subjective theory of welfare, one on which the subject's point of view on her life is authoritative for determining when that life is going well *for her*" (Ibid, 160, italics his). In this passage, one straightforward reading is that an agent's judgment is about, at least to some extent, her well-being.

There is a danger of vicious circularity with option (a). For with construal (a), the agent in the counterfactual situation is to judge whether some bit of happiness is or is not a constituent of his well-being. For example, Sumner says: “The extent to which the illusoriness of the experiences *matters* for an individual’s well-being therefore depends on the extent to which she decides (or would decide) to *make it matter*” (Ibid, pg. 161). If we fill out the expression ‘make it matter’ with ‘make it matter for well-being,’ the danger of circularity become clearer. The decision of the agent is supposed to provide us with a criterion for well-being, and yet the agent’s decision is about whether something matters for his well-being or not. I do not think that (b) or (c) lead to vicious circularity, and therefore I think them more charitable interpretations of what Sumner can have in mind.

Let us explore (b). With option (b), an agent’s endorsement merely consists in his remaining satisfied in a certain way after being relevantly informed. No passages in Sumner’s book support reading (b). Sumner simply does not say anything similar to (b). Nevertheless, (b) seems plausible to me and it is worth exploring in its own right. Yet I think that there is a general problem with Sumner’s mixed theory, even if (b) is the best way to go in explaining the counterfactual requirement.

I think that (c) is the best interpretation of Sumner’s position. Although there are a few ways of fleshing out the details of (c), all interpretations along the lines of (c) share the idea that endorsement of the sort Sumner has in mind is a desire of a certain sort—a desire for the conditions of one’s life. Endorsement can easily be understood as a desire, on my broad understanding of desires. I think of desires as pro-attitudes generally speaking. One can desire the conditions of one’s life. Nothing about desiring a state of

affairs, I think, requires that the state of affairs not already be the case. One might think of a desire for an existing state of affairs as a sort of valuing. Valuing of this sort could very well be an endorsement of the conditions of one's life.

There are some passages that seem to support interpretation (c). For example, Sumner says: "In order for a subject's endorsement of her life to accurately reflect her own priorities, her own point of view—in order for it to be truly *hers*—it must be authentic, which in turn requires that it be informed" (Ibid, pg. 160, italics his). Sumner's use of "priorities" here might be the key to understanding what he has in mind. Perhaps what Sumner has in mind is that one endorses the conditions of her life if she endorses her life satisfaction, even when the grounds of her life satisfaction turn out to be faulty.

Let me at this point take the opportunity to explain one key difference between Sumner's account and counterfactual desire-satisfaction accounts as I see them. Sumner's overall account is not *entirely* counterfactual. That is, his view is not that something contributes to your well-being, if it *would* make you happy under the right information conditions. Instead, Sumner's theory requires that one's life satisfaction to be actual and then requires that it pass the counterfactual condition, of which I think that (b) and (c) are two possible interpretations.¹⁴ Later, I will explore a weakness in Sumner's account that, I think, pushes Sumner to accepting a fully counterfactual account.

I think that one can offer a powerful argument in favor of pure desire-satisfaction accounts given what Sumner says about the endorsement condition—especially if we take (c) to be the most plausible interpretation of his view. Let me explain. Sumner says,

¹⁴ A desire satisfaction theory *could* be set up in a similar fashion. First, the actual agent must presently desire something. Secondly, the thing must still be desired were one to have full information about it.

variously, that one's well-being depends on her "priorities," "personal priorities," "point of view," "self-assessments," etc. However, one's priorities, self-assessments, etc., on option (c), are only relevant to one's well-being when there is actual life satisfaction in the picture. When there is no actual life satisfaction in the picture, one's priorities, point of view, self-assessments, etc. are irrelevant to her well-being. Sumner accuses theories other than his own of being unreasonably puritanical because they fail to give significant weight to a person's own priorities, yet Sumner's final account of well-being seems guilty of the same fault, given his line of reasoning. Sumner says "On a subjective theory, individuals are the ultimate authorities concerning their own welfare" (Ibid, pg. 171). Yet this is not so on Sumner's final mixed account. Even on option (c), Sumner's account requires actual life satisfaction to be in place in order for one to enjoy any degree of well-being. Here is the relevant question: *Why restrict the authority of the informed judge exclusively to matters of happiness?* On my account of well-being, the informed judge's opinion in matters of happiness is on par with his judgments about other elements that make up the person's life as to whether they matter for his well-being. On Sumner's theory, on this construal, only judgments on matters of life satisfaction are important. Of course, if we presuppose that the happiness theory is the best account of well-being, then the restriction is natural. But, such a presupposition is unwarranted and we are left wanting an argument.

Although interpreting the counterfactual condition in Sumner's theory as (c) results in a plausible theory and seems to have a good deal of support given what Sumner says, the arguments he at least seems to give in favor of his counterfactual condition really end up supporting a pure desire-satisfaction account, such as mine. Perhaps, given that

interpretation (a) results in a vicious circularity and that (c) does not fit well with Sumner's mixed theory, we should interpret him as having (b) in mind. However, he says very little, if anything, that supports interpretation (b). I have an independent argument against all mixed theories of the sort Sumner has in mind that I will offer later in this chapter. So even if we attribute (b) to Sumner, he does not escape serious criticism. At this point, however, I would like to finish exploring the conditions of Sumner's theory of well-being.

I will now discuss Sumner's account of the autonomy condition. The worry Sumner is trying to deal with comes from work by Amartya Sen. Sen thinks that happiness and desire-satisfaction theories are all subject to a criticism. Things like personal preferences and happiness are influenced by society, religion, family, friends and a host of other things. We can imagine cases where social conditioning might lead one to be happy or have a desire for a state of affairs, and yet one might feel that there is something "alien" about the happiness or preferences. A person could be happy and could have this happiness based on very good information and yet his happiness might not be "autonomous." A classic example of this in the philosophical literature is someone who is a slave and yet is content with his position, the feeling of contentment having been formed from social or other forms of conditioning.¹⁵ Sumner's way to get around counter-intuitive examples such as the contented slave is to require that the happiness be autonomous.

¹⁵ Another good example is from the Harry Potter series by J.K. Rowling. In those books, there is a race of what are called "house elves." House elves are, in effect, slaves to the humans who can use magic. Through the most recent book in the series (the sixth of what looks to be eight books), only one of the house elves, Dobby, has "thrown off the chains of oppression." The others are content to live as slaves. No doubt, what makes the house elves such good fictitious examples (and makes us think that they are "brainwashed" and not autonomous) is that most fiction that involves elves places them outdoors and communing with nature as opposed to indoors ironing clothes and washing dishes.

Sumner discusses two main accounts of autonomy: the identification/hierarchy view and the process/formation view. A desire or preference is autonomous on the identification/hierarchy view when the person has a higher-order desire in favor of it. This kind of autonomy requires critical reflection on one's own desires and preferences. Though much of Sumner's language on this topic is drawn from the preference and desire literature, his target of course is higher order desires about happiness. Sumner points out that the main problem with the identification/hierarchy conception of autonomy is that those higher-order values could themselves be socially influenced, and so would not solve the problem that faces his happiness theory (Sumner, pg. 167-171).

The process/formation view of autonomy is concerned with how a desire or preference is formed, not with whether it is identified with. The basic idea is that there are autonomy-subverting processes, such as indoctrination, programming, brainwashing, role scripting, etc (Ibid, pg. 171). Sumner points out the main problem with the process/formation view is that it fails to deal adequately with cases in which a goal is formed by an autonomy-subverting process and yet the agent later comes to identify with it.

Sumner concludes that both of the main conceptions of autonomy have problems and he does not adopt either in his autonomy condition. Still, he thinks that however the details of the autonomy condition are worked out, an autonomy condition needs to be included in his account of well-being. Because Sumner does not accept a particular conception of autonomy, it is hard to give precise criticisms of this idea. It is possible, I suppose, that no satisfactory account of autonomy will be developed. Also, it seems possible that, even if a satisfactory account of autonomy can be worked out, it might not

solve Sen's particular criticism. So there are some worries here, even though Sumner is confident that they can be addressed.

To summarize, Sumner's information requirement has many significant problems and his autonomy requirement is not spelled out in enough detail to enable us to know much about how it is supposed to work. I do have some other criticisms of Sumner's theory.

Sumner's theory has counter-intuitive implications that are easily explained by my desire-satisfaction theory. For example, we could imagine two people, *A* and *B*, who have the same amount of autonomous and informed life satisfaction and who are also equally misinformed about the grounds of their happiness. Let us suppose that they are equally deluded as to the actual state of affairs. If they were to have their mistaken beliefs corrected, then they would not endorse the conditions of their lives. Note that in discussing my example, I speak of Sumner's counterfactual condition in the vague or ambiguous way that Sumner does—I speak of “endorsing the conditions of one's life.” Both *A* and *B*, in my example, do *not* endorse the conditions of their lives. So far, everything I have said about *A* and *B* is the same. However, *A* and *B* differ: Were *A* to have his beliefs corrected, he would have life dis-satisfaction, but were *B* to have his beliefs correct, he would have neither life satisfaction nor dis-satisfaction. On Sumner's theory, both *A* and *B* enjoy the same amount of well-being—namely zero—because both *A* and *B* would not endorse the conditions of their lives. But that is an odd consequence *even when supposing that happiness, unhappiness, and information all matter for the theory*. Surely the correct outcome is that *A*'s life is going worse than *B*'s, though neither's life is going well.

The *pure* (mental state) happiness theory does horribly with the above example because it gives *A* and *B* both a high level of well-being. Sumner's *mixed* theory fares nearly as poorly because it gives both *A* and *B* an equally low level of well-being. These results seem implausible, even granting the supposition that happiness, unhappiness and information matter in a conception of welfare. This example reveals a central problem with Sumner's mixed theory, independently of which of the three construals of the counterfactual account I discuss above.

Of course, one could adopt a fully counterfactual view, according to which one's well-being is constituted by what would give one life satisfaction if one were informed in the proper way. Such a view might deal with the example I discuss above in an adequate way. One might be able to adopt a full counterfactual view and remain true to Sumner's insistence that information and endorsement matter. However, if one were to adopt a fully counterfactual life satisfaction account, such as the one just described, I believe that one would lose the original motivation for accepting at least a partially mental-state account in the first place. Surely, one of the central motivations for accepting a mental state account of well-being is that a mental state account countenances an essential relation between one's mental life and the level of one's well-being. On a fully counterfactual life satisfaction account, the connection between one's mental life and one's well-being would be lost. The fundamental nature of the account would be changed. Surely, desire-satisfaction accounts, such as my ideal advisor account, fare better than fully counterfactual mental state accounts. This is because desire-satisfaction accounts never depend on an essential link between one's mental life and one's well-being.

My example of *A* and *B* seems to force Sumner to face a dilemma. Either (1) he reverts to a pure mental state theory or (2) he adopts a fully counterfactual life-satisfaction account of the sort suggested above. With (1), Sumner would have to explain Nozick-style experience machine examples in a plausible way. This, it seems, he cannot do. At least, he does not seem optimistic that such an explanation can be given. Sumner was led to his mixed theory *because* of counterintuitive implications for pure mental state theories in experience machine problems. With (2), I say, Sumner would fare little better. Mental state accounts and counterfactual conditions do not mix well together because they have clashing motivations. A central motivation of mental state accounts is that they forge an essential connection between one's mental life and one's level of well-being. A central motivation of counterfactual information accounts seems to rest in the belief that one's mental states can be grounded in mistaken belief. In Sumner's mixed account there is a central tension between the rationale for the mental state component and the rationale for the full information component, and this makes it very unattractive. It seems that on either horn of the dilemma, Sumner faces serious difficulties.

6.3.2 Mental State Accounts Mixed with Objective Elements

Just as the desire theory can be mixed with objective components, so too can mental state conceptions. There is a large literature that treats Aristotle's supposed thesis that happiness is an activity in accordance with human excellence (*Nicomachean Ethics*, III and X.7-8). Some of it seems to suggest that Aristotle's account of *eudaimonia* is, in the terminology of this dissertation, a mixed account, according to which one has *eudaimonia* to the extent to which one has a certain kind of mental state such as happiness, provided that the happiness arises from the proper sorts of activities. The interpretation is controversial. More to the point here, my concern here is neither with

what happiness is nor with interpretive questions about Aristotle, such as whether he was indeed writing of happiness as at least partly a mental state. Of course there are many such interpretive questions, such as: “Is ‘*eudaimonia*’ best translated as ‘happiness’?” and “Is *eudaimonia* the same as the narrow conception of well-being I am after?” But these questions are off my topic. The parts of Aristotle’s theory that we can use to construct a mixed theory have already been developed in the earlier chapters. So I can side-step historical questions and questions about the nature of happiness.

A mixed theory that combines happiness with objective components might be motivated by the thought that happiness should not count toward a person’s well-being if it is grounded in morally unacceptable activity. People may be happy when torturing others or take pleasure in others’ suffering. But we cannot place moral restrictions on which bits of happiness contribute to well-being, for we are after a non-moral conception of well-being. So, moral considerations should not be taken into account. But a mixed happiness theory might also be motivated by the worry that happiness can be grounded in activities or thoughts that seem irrelevant to a person’s life. Instead of imagining a person having a plan or desire to count blades of grass or turn radios on and off, we can imagine a person who gets happiness out of such activities. Sumner’s focus on the type of happiness that is satisfaction with one’s life helps because it does exclude happiness that stems from judgments about how the lives of *others* are going; we might think of this as happiness got through vicarious living. However, I do not think that the focus on life satisfaction deals away with such examples completely. Nothing excludes a person from getting life satisfaction from turning on and off radios or counting blades of grass. So it may be that adding an objective requirement can help with these sorts of cases.

Let us look first at a combination of a perfectionist theory with a happiness theory. On this approach, an amount of happiness would contribute to a person's well-being only if the happiness were based in some relevant way in the development of human nature. Roughly, an account of this kind would be of the following form: A person enjoys well-being (to a certain degree) if and only if and because (i) the person enjoys happiness of a certain sort (and to that degree) and (ii) this happiness is based in the development of human nature (perhaps, to that degree). Call this theory "Mixed Perfection/Happiness" or "MPH" for short. Counting blades of grass or turning radios on would not accord with one's human nature, at least so a proponent of MPH could argue.¹⁶ If not, then happiness grounded in such activities would not contribute toward a person's well-being.

The familiar problems arise. An evolutionary account of human nature is attractive. Yet on an evolutionary account of human nature, the perfectionist requirement cannot significantly help mental state theories. There will be dispositions that are the result of natural selection, the pursuit of which do not increase one's well-being at all—in fact in some instances, there will be a selective (genetic) advantage to self-sacrifice.¹⁷

Perhaps the evolutionary account of human nature is wrong. Hurka thinks that human nature involves, roughly, having a physical body combined with two forms of

¹⁶ Sumner does consider a mixed theory just like this, though he calls it a hybrid theory. He thinks that the additional objective element would have to be a value requirement, somewhat analogous to the reality requirement. According to Sumner: "A value requirement, however, is even more questionable than a reality requirement, since it presupposes that there is an evaluative analogue to empirical truth or reality: a right answer to every question about value" (Ibid, pg. 164). In Sumner's mind, the value requirement would be objectionable for reasons similar to why the reality requirement was rejected: that it is unreasonably puritanical. I do not share Sumner's view here because I do not see the additional objective requirement as being evaluative: perhaps some may try to bring evaluation into the picture with a perfectionist requirement, but it is not the case that all would do so. This is especially the case when the objective requirement is brought in to solve issues of irrelevance, and not brought in for reasons stemming from moral or aesthetic values.

¹⁷ Please see Chapter 2 and the discussion of the mixed desire/perfectionist accounts found earlier in this chapter.

rationality, theoretical and practical. Development and exercise of one's physical body and of one's rationality are in one's self-interest.¹⁸ I presented criticisms of Hurka's theory in Chapter 2 of this dissertation and they are serious and still in effect. We need to hear more, for example, about what is included in the three types of excellences that we can exercise and develop and how they are to be made commensurate.

Turning radios on and off and counting blades of grass seem to involve *exercising* one's rationality, though it is less plausible that engaging in such activities *develops* one's rationality, except perhaps in those learning to count.¹⁹ Also, there are ways of exercising and developing the physical body that are mundane at best—for example Frankenstein the monster exercises his excellent digestive tract by eating grass and weeds. So unless I hear more about what these three conditions amount to, I do not think that they would necessarily solve the original problem of irrelevant happiness; in an intuitively plausible way.

Also, at least one new problem seems to crop up with the combination of the happiness theory and perfectionism in MHP. That is, MHP might exclude too much. It depends on what exactly we take happiness to be – on what the relevant mental state is. For now, following Sumner, let us say that the relevant mental state is life satisfaction. For a huge percentage of the population, it is hard to think of examples in which a person has life satisfaction that is based in activities that develop theoretical rationality. Most

¹⁸ Strictly speaking, Hurka does not take himself to provide an account of well-being, because he thinks that well-being is an essentially subjective concept. I disagree with that claim—my construal of the concept of well-being is neutral with regard to the so-called subjective and objective conceptions. So I can interpret Hurka's theory as concerned with well-being.

¹⁹ Perhaps it is the difficulties in spelling out these conditions that make Sumner think that the perfectionist requirement is a requirement of value: that there is not a value-neutral way of fleshing out three elements. That is just a guess as to what is lying behind Sumner's arguments, but it might nevertheless be accurate. I am not sure what to make of it.

people do not derive life satisfaction from doing complex mathematics or physics or similar things, not even if they do them well. Paradigmatic ways of exercising and of developing theoretical rationality will not enhance well-being when there is no properly corresponding life satisfaction—which I suspect would be the case for most people.

Someone who endorses the mixed theory under consideration might say “so much the worse for the general populace,” but that response should make us pause. One of the reasons for adding a happiness condition to the perfectionist picture is that the perfectionist theory seemed too austere, stipulating that there is one best way to live for everyone, or at least for each person, without any of it being *up to the agent*, in a way of speaking. It seems that the charge of austerity now arises again if the answer is “so much the worse for the general populace.”

Let us now look at mixed theories that combine a happiness theory with an objective list theory. I do not know of any one who has held such a view, so my discussion of it will be brief. We can combine the objective list theory and the happiness theory by requiring that in order for there to be an increase in a person’s well-being, there must be an increase in the person’s level of happiness and the increase in the person’s level of happiness must be based on some item on the list. For example, if aesthetic appreciation is on the list and if some agent is made happy by aesthetic appreciation, then there is an increase in the agent’s well-being. The mixed theory under discussion would be of the following form: A person enjoys well-being (to a certain degree) if and only if and because both (i) the person enjoys happiness of the required sort (to that degree) and (ii) this happiness is based in an activity of a kind that is on the list. Call this theory “Mixed Happiness/Objective List” or “MHO” for short. Because we can make up the list

anyway we want, MHO should solve any problems of irrelevant happiness—we can make up the list in such a way that simply excludes whatever seems irrelevant. Whereas it looks like MHP might not solve the problem of irrelevant happiness that afflicted pure happiness theories, MHO seems to solve it in an attractive way.

The devastating problem which completely undermines the objective list theories is incommensurability. Either the addition of the happiness requirement solves the problem of incommensurability or it does not.²⁰ If the new requirement does *not* solve the problem of incommensurability, then the account is still deeply troubled. Nevertheless, I suppose there is something to the idea that MHO is more plausible than a pure objective list theory, at least because of the element of “personal choice” does come into play through the subjective element in happiness. But we still do not know, for example, whether the slightest bit of happiness due to aesthetic appreciation increases well-being more than a huge amount of happiness due to friendship. So there is still a big problem for MHO if we cannot solve the incommensurability problem.

If adding the element of happiness does solve the incommensurability problem in objective list theories, then there is the possibility that the new mixed theory is plausible—at least it gets off the ground. The happiness condition can perhaps solve the problem on the basis of variations in the “strength” of happiness.²¹ Strength of happiness could solve the problem of incommensurability. If a person’s ground for a bit of

²⁰ The dilemma that faces the happiness/objective list conception is analogous to the one that faces the desire-satisfaction/objective list conception.

²¹ Following Mill’s thoughts on pleasure; there could be both quantity and quality of pleasures. Quantity consists in intensity and duration of a pleasure, while quality of pleasure comes by choosing one pleasure over the other—perhaps by choosing one pleasure over another when given the same intensity and duration, or some other scheme of choice. At first glance at least, this sort of quantity and quality scheme could be applied in a happiness theory. Call the overall scheme a “strength of happiness” scheme.

happiness is on the objective list, then the “strength” of that bit of happiness determines how much well-being is increased.²²

Despite the fact that MHO has some attractive features, it nevertheless is subject to a fatal flaw.²³ Once we adopt strength of happiness as determining the amount by which well-being is increased, the account is destabilized. We allow strength of happiness to determine the amount by which well-being is increased provided that the basis of the experienced happiness is on the list. There is no increase when the basis of the happiness is not on the list—the strength of an experience of happiness does not matter if its basis is not on the list. The natural question that arises is: “If strength of happiness matters in cases where a person has an experience of happiness and the grounds of the experience are on the list, why does it make no difference at all when the grounds of an experience of happiness are not on the list?” Heretofore, I have not required that the objective list have any justification, but when strength of desire is included to solve the problem of incommensurability, it brings the incompleteness of the objective list theory into clearer light. For unless there is some explanation for the list, then we cannot answer the question “Why does the strength of an experience of happiness matter in cases where the grounds of the experience are on the list, but not matter at all in cases where the grounds of the experience are off the list?”

²² I do not know if “strength” is the best word in this context. I speak loosely here. The idea is that each state of happiness has a certain degree of “strength.” Perhaps it would be better to speak of happiness as having intensities, or durations, or some other manner of degree. Nothing in my criticism of MHO hinges on a specific construal of what the “degree” of happiness consist in. I will use “strength” to capture the idea I have in mind.

²³ The problem with MHO is structurally similar to the core problem with MDR (Mixed Desire-Satisfaction/Rational Aim) I discuss in 6.2.1 of this chapter.

There is the general problem that we have no explanation for the objective list, but when a list view is supplemented by either strength of desire or strength of happiness to solve the problem of incommensurability, the problem becomes more pressing. The addition of the “strength of-” clause destabilizes an objective list theory. When the strength of a mental state such as happiness or desire *determines* the contribution to well-being of items on the list, we need an explanation of why the strength of such a mental state does not determine a thing’s contribution to well-being even if the thing is not on the list. So far no mixed theory that combines an objective list with happiness or desire-satisfaction has explained this.

6.4 Conclusion

I have examined several mixed theories and none of them does better than the pure desire-satisfaction theory. The mixed theories that involve a desire-satisfaction component turn out not to solve the original problems that the pure theories face or they turn out to have new problems of their own that arise from combining the desire-satisfaction component with the other component in the mixed theory. The same holds for mixed theories that include a mental state condition, such as Sumner’s theory. Sumner’s theory is perhaps the best developed and most plausible of the alternatives to ideal advisor theories, but Sumner’s theory turns out to have a host of significant problems; both in coming up with a plausible partial counterfactual condition and a plausible way to combine the mental state of “life satisfaction” with states of the external world.

CHAPTER 7

DO POSTHUMOUS EVENTS ALTER THE LEVEL OF WELL-BEING ONE HAD WHEN ALIVE?

7.1 Introduction

My conclusion of this chapter is that posthumous events do not alter the level of well-being that one had when one was alive.¹ However, I must do some groundwork before I get into the arguments for my conclusion. The issue of whether posthumous events change one's well-being arises from the prospectivity of desires. Let me illustrate: imagine something is desired at t_1 ; if the thing comes about at t_2 , and if the thing is still desired at t_2 , then the desire would be satisfied and there would be an increase in well-being. If one dies in between times t_1 and t_2 , then there is a question as to whether there can be an increase in one's well-being.

One assumption which will be made in this chapter is that there is not a substantial aspect or part of a person which remains after he dies. In this chapter, I wish to explore the implications of the desire-satisfaction theory on the possibility that no aspect of the person or personality survives the death of the person.²

Now, I do not explore whether there can be increases in the well-being one has after one is dead (in a sense that would require that there be an amount of well-being

¹ This chapter is modeled on a paper presented at the 2003 Ohio Philosophical Association. I would like to thank the audience there for many helpful comments.

² My arguments and conclusion in this chapter do have a number of implications for the view that only a part of the personality survives death—implications for those desires which are lost at death. However, I must restrict myself to exploring the possibility that none of the desire set survive death, given space constraints. Philosophers and theologians differ considerably on the extent to which the personality survives death. This dissertation has the most direct implications for the view that nothing survives death, but there still are important consequences for those accounts which maintain that some desires are lost.

without there being any remaining personality), for that is assumed not to be the case. To claim that the satisfaction of posthumous desires can increase well-being is to claim that events that happen after one dies might be able to affect the level of well-being one had when one was living.

Philosophers are pretty much split on the issue. Here is a short list of philosophers who think that events that happen after one's life could affect one's well-being, if a desire-satisfaction conception of well-being were correct: Brandt, *A Theory of the Good and the Right* (1979), Parfit, *Reasons and Persons* (1984), Griffin, *Well-Being* (1986), Hooker, "A Breakthrough in the Desire Theory of Welfare" (1993). On the other side are: Sumner, *Welfare, Happiness and Ethics* (1996), Overvold, "Self-Interest and the Concept of Self-Sacrifice" (1980), Fuchs, "Posthumous Satisfaction and the Concept of Individual Welfare" (1993). My conclusion is that posthumous events cannot change one's well-being. A desire-satisfaction account of well-being can be restricted to exclude the possibility of posthumous events' changing one's well-being, I will explain below, but hopefully there can be some *motivation* for whether the restriction should be made or not, so that such a restriction won't be *ad hoc*. Given the conflict of intuitions in this area, the best way to proceed is to look at ordinary, non-posthumous cases of desire-satisfaction, and generate principles that apply to posthumous cases.

7.2 Overvold's and Hooker's Proposals

One might be inclined to believe that the restriction of the relevant desire-set to those desires that are self-regarding would *determine* whether posthumous events can alter the well-being that one had when alive. If this were the case, then the debate over the relevance of posthumous events to well-being would be settled. In this part of the chapter, I will argue that the debate over the relevance of posthumous events is not settled

by my criterion of self-regarding desires. Let me first explain, however, why one might think that the debate over the relevance of posthumous desires could be settled by a criterion of self-regarding desires.

Mark Overvold thought that his proposed criterion of self-regarding desires entails that no posthumous event could alter one's well-being. Overvold, in his 1980 paper "Self-Interest and the Concept of Self-Sacrifice" is concerned with explaining how self-sacrifice is possible on a desire-satisfaction conception of well-being. Overvold is not primarily concerned with whether posthumous events can alter one's well-being.

As the reader may recall from an earlier discussion in this dissertation, Overvold's explanation of how self-sacrifice is possible on a desire-satisfaction conception is that, in cases of self-sacrifice, there will be a desire for some good other than the agent's own good, and we should think of this desire as being non-self-regarding. The satisfaction of a *non-self-regarding* desire will *not* increase well-being, nor will it decrease well-being. Only the satisfaction of self-regarding desires increases well-being. Thus, in cases of self-sacrifice, there is not any increase or decrease in the agent's well-being that results simply from the satisfaction of the intrinsic desire for whatever good the agent has in mind since that desire is non-self-regarding. The decrease in the agent's well-being that occurs in cases of self-sacrifice comes from the dis-satisfaction of other intrinsic desires, desires that are self-regarding.

Overvold thought that to exclude non-self-regarding desires from the relevant desire set (the satisfaction of which increases one's well-being), it is necessary to add the following condition: A desire for a feature or outcome is a member of the relevant desire set only if "the proposition that the agent exists at t is a logically necessary condition of

the proposition asserting that the feature or outcome obtains at *t*” (Overvold, pg. 118).

Another way in which Overvold describes this condition is that the agent must be an “essential constituent” in the desired outcome or state of affairs (Ibid, pg. 118). Overvold suggests another condition—that the *reason* for the desire be self-regarding, in addition to the “essential constituent” condition, but given the topic of this chapter, this second condition is irrelevant.

Were Overvold to be right in his account of the distinction between self- and non-self-regarding desires, then what he says would provide an independent criterion for thinking that posthumous events do not alter the well-being one had when one was alive. As the reader may recall, however, I think that Overvold’s condition does not work to make a clear distinction between self- and non-self-regarding desires. This much, I hope is clear. However, things are not as simple as they seem. Let me explain.

As I briefly mentioned in 5.1.4, Brad Hooker, in his “A Breakthrough in the Desire Theory of Welfare,” proposes a way of altering Overvold’s account so as to include the possibility of posthumous events altering the level of well-being one had when one was alive. As I will explain, Hooker’s proposed alteration can be clarified and adjusted to fit my account of the distinction between self- and non-self-regarding desires.

Hooker states:

We might modify this [Overvold’s proposed necessary condition] so that the relevant desires are the ones in whose propositional content the agent is an essential constituent in the sense that the state of affairs *is desired under a description that makes essential reference to the agent*. By this criterion, desires such as the desire that *I* be influential (or posthumously famous) or the desire that *I* have been a good parent count, though events after my death can determine whether or not such desires will be fulfilled. (Hooker, pg. 212, emphasis his).

Now, I do not think that Hooker is as clear as he could be in this passage. Unfortunately, he provides no further explanation of his proposed modification. Nevertheless, I think

that a proposal along the lines of Hooker's can be made to fit my theory. Let me explore an adaptation of Hooker's proposal to fit my proposed criterion of self-regarding interests.

Hooker seems to have in mind that, instead of requiring that the agent be a constituent of the desired state of affairs, we should merely require that the state of affairs is desired under a description that makes essential reference to the desiring agent. Now, presumably what Hooker has in mind is that the range of relevant desire descriptions is fixed by the desiring agent's beliefs. Let me explain. Imagine that Lois Lane has the desire that she kiss Superman. Someone who, unlike Lois Lane, knows that Superman is the same person as Clark Kent, might describe Lois Lane's desire as a desire to kiss Clark Kent. Yet Lane would never describe her desire as one to kiss Clark Kent. What Hooker might have in mind is, to be precise: in order for an agent's desire to be self-regarding, the agent must be an essential constituent in the desired state of affairs as the agent would describe it from the agent's informational perspective.

My proposed distinction between self- and other-regarding desires is a bit more sophisticated than Overvold's. Recall that I claim in Chapter 5 that if a sentence aptly expressing a desire makes reference to the relevant agent, *A*, in the "object sentence," then the desire is *A*-regarding. An "object sentence," recall, is a sentence aptly expressing the desired state of affairs. An "object sentence" is to be distinguished from a "subject clause." A "subject clause" attributes a desire to an agent. Reference to *A* in a "subject clause," whether *de dicto* or *de re*, is neither necessary nor sufficient for a desire to be *A*-regarding.

Hooker's proposal, as I understand it, could be a restriction on the relevant desire set, even on my proposal for the distinction between self- and non-self-regarding desires. Hooker's proposal, as I understand it, when adapted to my theory would be that if A is referred to in a sentence aptly expressing the desired state of affairs of desire, d , under a description that is appropriate given an agent's informational perspective, then d is A -regarding.

I suspect that the above proposal could be made to work. I understand that I have left out some details of how the proposal is to work. What I take myself to have shown is *not* that a desire-satisfaction account of well-being *should* allow for posthumous events to alter one's well-being. Instead, I take myself to have provided a *prima facie* case that a desire-satisfaction account that uses my distinction between self- and non-self-regarding desires in the way I recommend can go either way on the "posthumous-events-altering-well-being" debate. That is, what I hope to have shown is that there is good reason to think that my ideal advisor account, even using my distinction between self- and non-self-regarding desires to identify the relevant desire set, is flexible enough to either count or not count posthumous events as being relevant to well-being. Rather than explore the details of the proposal in a more rigorous fashion, I wish to argue that it is unnecessary to provide the details of the account. The details are unnecessary, I will argue, because it is a mistake to think that posthumous events can alter well-being on desire-satisfaction accounts of well-being.

7.3 A New Beginning

We have to look at the issue anew. There really are two questions that are often conflated when thinking of posthumous events and the desire-satisfaction theory of well-being: (1) Can desires be satisfied when they are no longer had, and (2) If so, would the

satisfaction of desires after one has died yield increases of well-being? I answer a somewhat cautious “yes” to the first and a “no” to the second. As I understand it, saying “no” to either would be sufficient for rejecting posthumous increases in well-being. However, I should note that answering “yes” to (1), as I propose to do, has implications that should make us think that posthumous events cannot alter the well-being one had when one was alive – at least not on desire-satisfaction accounts.

The idea behind the following arguments is that we must have a principled way of determining when a desire is satisfied and when well-being is increased. My main arguments at this point rely on intuitions in cases where people are alive. The case of posthumous increases in well-being is difficult. My plan is to look to the easy cases first and then to generate principles that apply to all cases, including posthumous cases.

Here are some cases that make me think we should answer “yes” to (1). Imagine that there is someone who is a terrorist. He desires, at t_1 , that a ship sinks at t_2 . Now imagine at t_2 that he no longer has the desire or even desires that the ship not sink. Furthermore, imagine that at t_2 the ship, in fact, sinks. Are we to say that the desire he had at t_1 is satisfied? This is a very difficult question to answer and I think most people will have conflicting intuitions about it. Here is another case that should help.

Imagine now that someone is a terrorist and desires that he sink a ship on a suicide mission (please take it for granted that the terrorist’s desire that he sink a ship on a suicide mission is intrinsic and sufficiently self-regarding). He will die in this mission but imagine that he ultimately accomplishes his goals. His death occurs just before the ship sinks, even if just by a split second. It would be odd to say to him, if he is considering whether or not to sink the ship, that his desire could not be satisfied because

he will die and will thus not have any desires. Of course, this is not conclusive, but it looks now like the answer to the first question about whether a desire can be satisfied when it is no longer had is “yes.”

Let us look a bit closer at what it is to attribute a desire to a person at a time and what it is for a desire to be satisfied. In the first case discussed, the terrorist does not desire the state of affairs to obtain, and yet it does anyway. When we attribute desires, we do so to a particular person and at a particular time. If he does not have the desire to sink the ship anymore, it would be misleading to say that it is *his* desire that he sink the ship. Thus, we cannot say that *his* desire is satisfied. But we can say that the desire he had is now satisfied. The statement ‘there is *a* desire which no one now has which is satisfied’ strikes me as confused. But it is not implied by the statement ‘a desire he had is now satisfied.’ The basic lesson we can learn from this subtler, more considered analysis, is that a desire does not have to be attributed to an agent, at a time, for it to be satisfied, at that time.

In the second case, the case of the suicide mission, a similar response is appropriate. A desire that the suicide bomber had is satisfied. The satisfaction does not occur when the state of affairs is desired, so properly speaking, the desire is not one that he has when it is satisfied. Still, the suicide bomber says, when deliberating about whether to sink the ship, “my desire that I have now will be satisfied then, even if I no longer concurrently have it” and that is perfectly intelligible. I think that the basic lesson mentioned in the first case is taught by this second case as well, *a desire does not have to be concurrently held and satisfied for it to be satisfied*.³

³ Numerous works by Brandt also argue this same point, see his “Overvold on Self-Interest and Self-Sacrifice,” among other works.

The reader still might feel that there is something metaphysically suspicious in the statement that “a desire that he had is now satisfied.” Here is an analogy with belief that might help make things less mysterious. Suppose someone believes a proposition, p , at time t_1 . The person ceases to believe p at t_2 and thereafter. Imagine that p is true at some time after t_2 : call this time t_3 . If t_3 is now, then a belief he had is now true. This statement of truth in belief has the very same structure as the case of desire discussed above. So although there might initially seem to be something mysterious about the claim that a desire does not have to be concurrently held and satisfied, it is perfectly intelligible, so long as one is careful about the tenses of the sentences used to describe such cases.

This basic lesson has important implications if we are to treat well-being as desire-satisfaction. We cannot just allow any desire satisfaction to count in favor of one’s well-being if a desire can be satisfied when it is no longer held. This makes me think that we must place a principled requirement on which desires are such that satisfaction of them can count towards one’s well-being. I suggest that the requirement is that the desired state of affairs must occur while the agent still has the desire for the state of affairs in order for satisfaction of the desire to increase the agent’s well-being. So in order for well-being to increase, it must be the case that both the person desires the state of affairs at time t and that the state of affairs obtains at that same time t . Call this principle, the “Concurrence of Desires and Obtainings Principle.” This principle allows that a desire may be satisfied by events that occur after it no longer exists, but it denies that such satisfactions increase well-being. The principle places a restriction on which desire-satisfactions lead to increases in well-being.

The Concurrence of Desires and Obtainings Principle allows for both cases in which the desired state of affairs comes about when the desire is held, and cases in which the state of affairs already is the case when the desire is formed. Both options are fine. The required concurrence is between the desired state of affairs and the desire.

People's desires change over time and well-being is determined by the extent to which their present desires match up to the present world, I say. The alternative of allowing past or future desires (either in exclusion of, or in addition to, present desires) to be relevant to present well-being is not plausible. It would be quite implausible to allow only past or only future desires to count as relevant to well-being. It would be barely more plausible to allow past, present and future desires to count as relevant to well-being. Let us call desire-satisfaction accounts of well-being that treat past, present and future desires as relevant to one's well-being "pan-temporal desire theories." "Pan-temporal desire theories" have extremely counter-intuitive implications. Let me explain.

Firstly, pan-temporal theorist would need an account of when and to what extent one's past, present and future desires are relevant to one's well-being. Such a principle, it seems, would have to be extremely complex. Perhaps pan-temporalists would consider satisfaction of present desires to enhance well-being to a greater extent than past and future desires, but perhaps not. Additionally, it is not clear to what extent the length of time one has a desire is relevant to one's well-being.⁴ The principle would have to be very complex indeed. This is especially so in the following sort of case: imagine that at t_1 A desires p . Later, at t_2 A desires not p . I think that A 's change of heart, if the details were fleshed out, could be a simple and ordinary event—something that happens all of

⁴ Note that the length of time having a desire is distinct from the "strength" of desire that all desire-satisfaction accounts must explain.

the time in ordinary life. Now, it is far from clear how pan-temporal theories would explain such a case. Desire-satisfaction theorists who reject pan-temporalism, such as myself, explain the case in a simple way: the obtaining of states of affairs that makes p true enhances A 's well-being at t_1 , but not t_2 . I do not know what principle pan-temporal theories would have for explaining cases of changing desire. Perhaps, pan-temporal accounts will treat both of A 's desires as relevant to his well-being at any given time and they will "cancel each other out." Perhaps pan-temporal accounts will allow present desires to "trump" past desires when the two are for conflicting states of affairs, though seems *ad hoc*. I am far from certain that pan-temporal desire theories could have a plausible principle for explaining problem-cases of the sort just described.

Secondly, it seems that pan-temporal desire theories do not deal well with changing personalities. I have suggested that one of the motivations for accepting desire-satisfaction theories, and more particularly, ideal advisor theories, is that they treat one's personality as relevant to one's well-being. The relevant desire set must reflect the full range of one's personality and desires in the relevant set cannot distort one's personality, I say. Now, one's personality can change over time. Given this fact, A 's personality at t_1 must be reflected in his relevant desire set at t_1 . If A 's personality at t_2 is different than it was at t_1 , then A 's relevant desire set must reflect his new personality. Pan-temporal desire theories fail to allow for the relevance of personality in an intuitive way. This is not a conclusive refutation of pan-temporalism, I think, because pan-temporalists might be able to make the required changes in order to have a plausible theory. But such changes would, I think, be *ad hoc*.

I have offered two arguments that each offer good reasons to reject pan-temporal desire theories. I maintain that the present character and personality of the person, the present desire set, is what is relevant for well-being.

If I am right, then posthumous increases in well-being are impossible. This is because once someone is dead, on the notion taken for granted in this essay that no part of the person survives death, he no longer desires things. After death, there is no possibility of desires and states of affairs concurring. Even if the state of affairs that was desired comes about, this still is not a case of desire satisfaction that increases well-being, for the desire and its satisfaction are not concurrent.

If I am wrong in my answer to (1) above, then there cannot be posthumous desire satisfactions and hence there cannot be posthumous increases in well-being on a desire-satisfaction view. But suppose I am right—there can be posthumous desire satisfactions. Suppose furthermore, that I am wrong about the Concurrence of Desires and Obtainings Principle and that we should reject this principle. If all of this were supposed, we would still have to settle on an account of which desire-satisfactions entail increases in well-being. Recall that I claimed that no increase in well-being can occur after one is dead. There cannot be any well-being after one is dead, much less an increase in well-being. However, if we reject the Concurrence of Desires and Obtainings Principle, we need to return to the idea that the occurrence of a desired state of affairs may *retroactively* increase the well-being one had when one was alive. There are several problems with this idea.

First, anyone who wishes to claim that desires which are satisfied after one has died can increase one's well-being must face a challenge about causation. The possibility of

retroactive *causation* is remote at best. Indeed, retroactive causation seems impossible. The challenge is to explain the nature of the relation between desired posthumous events and the level of well-being one had when alive without making use of the idea of retroactive causation – or else (the dimmer prospect) to explain how retroactive causation would work. I do not have any settled ideas on how either explanation would go.⁵ So there is a challenge to anyone wishing to adopt the thesis that posthumous desire-satisfactions can alter the well-being the agent had when the agent was alive.

Second, leaving the challenge about causation aside, there is yet a further hurdle to overcome. There must be a principle that tells us when the increase in well-being occurs, given that a desired state of affairs obtains. My hypothesis is that the increase in well-being must occur at the time when the desired state of affairs comes about (or as soon as the desire for it is present). Here is a hypothetical case similar to one mentioned earlier: imagine that a person makes it her project to rehabilitate a run-down city park. Imagine furthermore, that at a later time she accomplishes her goal and does so when she is still alive. When does the increase in well-being come about: when she accomplishes her goal, or some other time? The intuitive answer is that the increase comes when she makes the city park nice again—when she accomplishes her goal. We do not think, in the case where people are alive at the time of the accomplishment of their goals, that the increase in well-being is retro-active. There are other views one might take about when the increase in well-being occurs, but they are implausible; for example, there is the idea

⁵ One interesting possibility that has come up in discussion is to try to make use of the concept of supervenience. When the relation of supervenience holds between two properties (say, between the property of having a certain level of well-being and the property of having some desires be satisfied), the subvening property might not *cause* the supervening property. Perhaps the supervening property can be thought of as *constituted*, but not caused, by the subvening property. This, however, does not seem to make things much clearer because now one must give an explanation of retroactive supervenience—which seems just as mysterious as retroactive causation.

that the agent's level of well-being increases when the desire is formed or the idea that his level of well-being increases steadily between the formation of the desire and the obtaining of the goal. These proposals have the unfortunate feature that they imagine an increase in well-being when nothing in the world has yet changed to bring about the satisfaction of the desire. So, the most plausible view is that an increase in well-being occurs at the time when the desire is satisfied. Call this second principle the "Co-Active Increases Principle." It rules out retroactive increases in well-being.

If I am right about the Co-Active Increases Principle, then posthumous increases in well-being are not possible. If an increase in well-being can only come when the desired state of affairs comes about, and if the person has died by the time the state of affairs comes about, then there cannot be any increase in well-being. This is because a dead person has no level of well-being at all. The person who thinks that posthumous desire-satisfaction will lead to increases in well-being must deny the Co-Active Increases Principle and must suppose that there can be retro-active increases in well-being. Again, this is because it would be impossible to get an increase in well-being at a time when one is already dead. I argue that there are no such retro-active increases in well-being. For I argue that we should accept the Co-Active Increases Principle.

The process used in this chapter has been largely one of reflective equilibrium—mediating between principles and particular hypothetical cases in determining the best overall features of the desire-satisfaction theory and exploring the prospects for the view that posthumous events can alter one's well-being. Although the principles are clear, there may be some remaining intuitions which oppose them. There is, however, an

explanation that vindicates these intuitions while remains true to what has been argued in this chapter.

Let us consider some examples of cases in which one might reasonably think that events which happen after one's death alter one's well-being. Perhaps the best examples are examples of great artists who struggled throughout their lives in obscurity only to become famous when they were dead. Two prominent cases are Vincent van Gogh and John Kennedy Toole.⁶ Their desires to create masterpieces were indeed satisfied when they were alive. But if they desired to be *recognized* as great artists, then this only happened after their deaths. We might think that they gained in well-being as a result of becoming recognized even if this happened after their deaths. Now I can agree that the obtaining of states of affairs that one desired when one was alive might make for a better life in some sense, although not by increasing one's level of well-being. For example, in the cases of our artists, perhaps posthumous recognition led to increases in the moral quality of their lives, and especially to increases in the aesthetic or literary quality of their lives. A life story often has a better biographical quality when the things we wanted when alive come about, even after they come about after we have died. The narrative relations between events that make up each of our biographies may very well include things that happen before we are alive and things which happen after we are dead. So I can explain, perhaps, the genuine root of our intuitions without giving up the view I have been arguing for. To fully explain our intuitions, however, and to vindicate my account

⁶ Van Gogh and Toole received very little recognition as great artists during their life times. Van Gogh sold only one of his paintings and lived in nearly constant poverty. Toole is certainly less well-known than Van Gogh, but their basic life-stories are somewhat similar. Toole wrote *A Confederacy of Dunces* which was rejected for publication. He, like Van Gogh, committed suicide. Toole received the Pulitzer Prize posthumously for his novel.

of what lies at their root, I would need to say more to explain the aesthetic value or the narrative structure of a life story (which could include events before and after life). Nevertheless, the moral value and aesthetic value of a life are distinct from the non-moral prudential value of well-being.

I have tried to argue that there are not posthumous increases in well-being (granted the assumptions in the early part of the chapter). I allow that there can be posthumous desire satisfactions. But I think that once we look into what counts as desire satisfaction, we must restrict the desire-satisfactions that are to count towards our well-being. This is where the “Concurrence of Desires and Obtainings Principle” comes in. The Concurrence Principle is a principle about *which* desire-satisfactions should count as relevant for well-being. Also, I deny that there can be posthumous increases in well-being because I think that an increase in well-being occurs at the time that a desire is satisfied. That is, I think that an increase in well-being can not be retro-active. This is where the “Co-Active Increases Principle” comes in. The Co-Active Principle is a principle about *when* a desire satisfaction leads to an increase in well-being. If either one of these principles is correct, posthumous events cannot alter well-being. Both are correct.

7.4 Concluding Remarks

In this dissertation, I have developed and defended a desire-satisfaction conception of prudential well-being. My theory is somewhat complex, so it makes sense briefly to re-state what I take myself to have done in this work.

Let me take this opportunity to present my ideal advisor theory with all of the developments I have made in this dissertation in place. For a person *A* and her ideal counterpart, *A+*: *A*’s well-being varies to the extent to which *A*’s life accords with *A+*’s

A-regarding, non-moral and non-aesthetic intrinsic advice to *A* for *A*'s sake. *A*⁺ is *A* as *A* is in the nearest possible world to *A*'s actual world in which *A* (1) shares a personality with *A*, (2) is informed, as if by miracle, with as much information as possible given the nature of his perspective and personality, and (3) cares for *A* as *A* is in *A*'s actual world.

In the first chapter, I attempt to explain, in an intuitive way, the concept of well-being that constitutes the subject matter of the rest of my dissertation. There seem to be at least two different concepts of well-being in the philosophical literature: a broad concept of well-being and a narrow concept of well-being. The broad concept seems to involve overall choiceworthiness in a life and could include, *inter alia*, moral considerations. A narrower concept of well-being involves merely prudential well-being. In the chapter, I attempted to clarify the concept of prudential well-being using several examples from *The Count of Monte Cristo*, *The Picture of Dorian Gray*, and other sources. Prudential well-being, I claim, consists in the satisfaction of one's non-moral, non-aesthetic, self-regarding interests. That is, this is the concept of prudential well-being, although of course the concept can be explicated in different ways. So, I say, there are many "conceptions" of prudential well-being, or many different ways of explicating the concept of prudential well-being.

In the second chapter, I begin to argue for a particular conception of well-being—one that treats well-being as consisting in desire-satisfaction. I discuss three monistic conceptions of well-being: (a) explanatory objective theories, (b) mental state theories, and (c) desire-satisfaction theories. Additionally, I discuss a few pluralistic conceptions of well-being.

Explanatory objective theories include human nature theories. I find them lacking on several fronts. I will briefly mention two problems. Firstly, I do not think that proponents of human nature accounts of well-being have done an adequate job in *disproving* the thesis that our human nature is given to us by an evolutionary process. If our nature is given to us by an evolutionary biological process, any human nature account of well-being would be implausible, because we are evolutionarily disposed to do many things that, intuitively, are not in our self-regarding interests. Secondly, even if an explanatory objectivist could get around the problems of specifying a non-biological account of human nature, a significant problem that remains is that once an explanatory objective account is fleshed out, the theory is not sufficiently self-regarding. I rely on examples from *Frankenstein* to make my argument.

Mental state theories generally treat well-being as consisting in happiness or pleasure. I find mental state theories inadequate because they fail to provide a plausible account of one's well-being when one's mental life is not relevantly grounded in facts about the "outside" world. I rely on Nozick's experience-machine example and others to make my case. There is, I think, a way for a mental state theorist to fix problems that follow from experience-machine cases. In particular, I think a mental state theorist can add a requirement that the relevant mental state be appropriately connected to the "outside" world. I put off discussing such "alloyed" theories until the sixth chapter.

Pluralistic theories of well-being treat at least two factors as each independently sufficient to constitute an enhancement of well-being. I admit that some pluralistic accounts of well-being, such as the views of James Griffin and John Finnis, do a good but imperfect job of explaining our intuitions in many cases. Without a principle of

commensuration, however, pluralistic theories have no way, even in principle, of giving a determinative “verdict” about one’s level of well-being when one has a certain level of two or more of the conditions which are each independently sufficient for an enhancement of well-being. I explore ways in which a pluralist might respond to my concerns and find none of them promising.

In Chapter 3, I develop my preferred desire-satisfaction account of well-being. I begin with a discussion of intrinsic and extrinsic desires. I argue that only the satisfaction of intrinsic desires can enhance well-being. My argument is based on a problem of “double-counting” for any view that allows the satisfaction of extrinsic desires to enhance well-being. Next, I explore the plausibility of a view—an actual intrinsic desire-satisfaction view—according to which one’s well-being is enhanced when one’s intrinsic and actual desires are satisfied. I find it to be implausible because of examples involving incomplete or incorrect information.

Concerns with an actual intrinsic desire-satisfaction view lead us to consider counterfactual desire-satisfaction accounts. A central motivation for counterfactual accounts, I think, is that desires in the relevant set must reflect one’s personality. Recall that by “relevant set” I mean “the set of desires the satisfaction of which increase one’s well-being.” I take Railton’s account of a personality, develop it, and explain what I take its role in desire-satisfaction accounts of well-being. If one has incomplete information, then one will have a desire set that does not appropriately reflect the full range of one’s personality. If one has incorrect information, then one’s desiring might distort one’s personality. Brandt adopts a counterfactual desire-satisfaction view of sorts, but I argue

that his theory falls prey to the same arguments that defeat the actual intrinsic desire-satisfaction view.

According to ideal observer views, one's well-being varies to the extent to which the desires are satisfied that one would have were one to be fully informed and reasoning correctly. Although I find ideal observer views quite plausible, they exclude any possibility of desires one has for acquiring *new* information. If one had a desire for acquiring new information, then that desire would be satisfied and then "extinguished" upon being fully informed. For example, perhaps a person who actually desires to get to know the taste of pineapple and then is fully informed would cease to desire to get to know the taste of pineapple. Ideal observer theories, though plausible quite generally, can be improved upon because of the problem of "extinguished" desires.

The view I endorse is an ideal advisor account. It deals nicely with the problem of "extinguished" desires. According to Railton's statement of an ideal advisor theory, one's life goes well to the extent to which (roughly) the desires one would want oneself to want, were one to be fully informed and to be reasoning correctly, are satisfied. I endorse an account that is slightly different, which I call a "streamlined" theory. A "streamlined" ideal advisor theory holds that one's well-being varies to the extent to which one's life accords with what one would want for oneself, were one to be fully informed and to be reasoning correctly. A's well-being on my account, varies with the extent to which A's life accords with what A+ wants for A, for A's sake, where A+ is just A as roughly he would be if he were fully informed. It is in virtue of the facts that A and A+ share a personality and that A+ is informed in the right way, I argue, that A+ is "authoritative" with respect to A's well-being.

In Chapter 4, I address several criticisms that have been leveled against ideal advisor accounts in recent years. I argue that ideal advisor theories, possibly along with virtually all accounts of well-being I discuss in my dissertation, meet a weak “internalism” requirement according to which, roughly, the fact that something would enhance a person’s well-being would motivate the person appropriately. Rosati advocates a stronger internalism requirement, which ideal advisor accounts do not satisfy, but her requirement lacks any significant motivation.

Given the counterfactual nature of my account, people might worry about problems stemming from the so-called “conditional fallacy.” I argue that such worries are ungrounded because I can, in principle, “pick out” the ideal advisor of *A* as *A* in the nearest possible world in which *A* is fully informed, shares the actual *A*’s personality, and so on.

Writers have suggested that no one could be fully informed in the way that ideal advisor accounts require. I admit that it might be impossible for one to be “fully” informed if we understand “full information” in a certain way, but I think that all that an ideal advisor account needs is that one’s ideal advisor be as fully informed as the nature of one’s personality and perspective allow. Additionally, I argue that even if one’s informed counterpart lacks some non-propositional knowledge, such as knowledge of what it would be like to live certain ways of life, this is not as problematic as some have supposed. Lastly, I discuss Railton’s views about the role of personality in ideal advisor theories of well-being.

In the fifth chapter, I take the ideal advisor account, as I have developed it to that point, and turn it into an account of prudential well-being. I argued in Chapter 1 that

prudential well-being consists in the satisfaction of one's non-moral, non-aesthetic, self-regarding interests. I am influenced by Mill to think that a desire-satisfaction account of prudential well-being is slightly more plausible than it would otherwise be if it excludes one's moral considerations from factoring into one's desires or preferences. I explore two ways of eliminating moral and aesthetic considerations from factoring into my ideal advisor account.

Virtually any desire-satisfaction account of well-being has to distinguish between desires that are agent-regarding and those that are not. I argue for a distinction that treats desires as *A*-regarding if and only if there is reference to *A* in any object sentence aptly expressing the desired state of affairs: as I explain in detail in the chapter, this is so for both *de re* and *de dicto* desire attributions. My way of characterizing the distinction between *A*-regarding and non-*A*-regarding desires makes, I think, for a clear and intuitively attractive way of partly identifying the set of desires the satisfaction of which increase one's well-being.

It does seem that, even with all of the developments made thus far, it would be possible for one's ideal advisor to want something for one that, intuitively, would not enhance one's well-being. To counter this concern, I claim that adding a requirement that one's ideal advisor care for one takes care of most of the problematic cases in an attractive way.

In the sixth chapter, I return to critical discussion of various conceptions of well-being that compete with my own, including especially "mixed" theories. I argue that no mixed theory is more plausible than my ideal advisor theory. Mixed theories are interesting in that they are often constructed in an attempt to solve problems with so-

called pure theories. Sumner's authentic life-satisfaction theory is perhaps the most plausible alternative to my own theory. Yet his theory has several problems—most notably with the counterfactual condition of authenticity. My ideal advisor theory is better than its competitors as an account of prudential well-being.

In Chapter 7, I argue that posthumous events cannot alter the level of well-being one had when alive. If desires cannot be satisfied when they are no longer had, then on a desire-satisfaction view, posthumous events cannot alter the level of well-being one had when alive. If desires can be satisfied when they are no longer had, I think that desire-satisfaction accounts of well-being are implausible unless they require that well-being increases only when a desire and the obtaining of its object are concurrent. This is a condition on the relevant desire set. Additionally, I argue for a “Co-Active Increases Principle.” This principle is that well-being increases only when one has a desire and the relevant state of affairs obtains. This is a condition on when well-being increases given the satisfaction of a member of the relevant desire set.

This concludes my defense and exposition of a desire-satisfaction conception of prudential well-being. The account I have defended is better described as an ideal-advisor account. I believe it is more plausible than any competing account.

LIST OF REFERENCES

- Aristotle (1980) *Nicomachean Ethics*. Translated by W.D. Ross. Revised by J.O. Urmson. Oxford, England: Oxford University Press.
- Austen, Jane (1815) *Emma*, in *Jane Austen: The Complete Novels*. (1981) New York, NY: Gramercy Books.
- Brandt, R.B. (1979) *A Theory of the Good and the Right*. Oxford, England: Clarendon Press of Oxford University Press.
- Brandt, R.B. (1993) "Overvold on Self-Interest and Self-Sacrifice" in John Heil (Ed.), *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold*. Lanham, MD: Rowman & Littlefield Publishers, Inc, pg. 221-323.
- Brink, D.O. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge, England: Cambridge University Press.
- Calloway, Stephen (1997) "Wilde and the Dandyism of the Senses" in Peter Raby (Ed.) *The Cambridge Companion to Oscar Wilde*. Cambridge, England: Cambridge University Press, pg. 34-56.
- Carson, T. (1993) "The Desire-Satisfaction Theory of Welfare: Overvold's Critique and Reformulation" in John Heil (Ed.), *Rationality, Morality and Self-Interest: Essays Honoring Mark Carl Overvold*. Lanham, MD: Rowman & Littlefield Publishers, Inc, pp. 233-246.
- Carson, T. (2000) *Value and the Good Life*, Notre Dame, IN: University of Notre Dame Press.
- Crisp, R., and Hooker, B. (2000) *Well-being and Morality: Essays in Honour of James Griffin*. Oxford, England: Clarendon Press of Oxford University Press.
- Chang, Ruth (1997) "Introduction" in her *Incommensurability, Incomparability, and Practical Reason*. Cambridge, MA: Harvard University Press, pg. 1-34.
- Darwall, S.L. (1997) "Reasons, Motives and the Demands of Morality: An Introduction" in *Moral Discourse and Practice: Some Philosophical Approaches*. Oxford, England: Oxford University Press, pg. 305-312.
- Darwall, S.L. (2002) *Welfare and Rational Care*. Princeton, NJ: Princeton University Press.

- Dumas, A. (1990) *The Count of Monte Cristo*. Oxford, England: Oxford University Press.
- Dworkin, Ronald (2000) *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press.
- Feinberg, Joel (1989) "Autonomy." In John Cristman (Ed.) *The Inner Citadel: Essays on Individual Autonomy*. Oxford, England: Oxford University Press, pg. 27-53.
- Finnis, J. (1980) *Natural Law and Natural Rights*. Oxford, England: Clarendon Press of Oxford University Press.
- Fuchs, A. (1993) "Posthumous Satisfaction and the Concept of Individual Welfare" in John Heil (Ed.), *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold*. Lanham, MD: Rowman & Littlefield Publishers, Inc, pp. 215-220.
- Gibbard, Alan (1989) *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Griffin, J. (1996) *Value Judgement: Improving our Ethical Beliefs*. Oxford, England: Clarendon Press of Oxford University Press.
- Griffin, J. (1986) *Well-being: Its Meaning, Measurement, and Moral Importance*. Oxford, England: Clarendon Press of Oxford University Press.
- Hart, H.L.A. (1961) *The Concept of Law*. Oxford, England: Oxford University Press.
- Heil, J. (Ed) (1993) *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold*. Lanham, MD: Rowman & Littlefield.
- Hooker, B. (1993) "A Breakthrough in the Desire Theory of Welfare" in John Heil (Ed.), *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold*. Lanham, MD: Rowman & Littlefield Publishers, Inc, pp. 205-214
- Hurka, T. (1996) *Perfectionism*. Oxford, England: Oxford University Press.
- Kagan, S. (1992) "The Limits of Well-Being" in Ellen Paul, Jeffrey Paul, and Fred D. Miller (Eds.) *The Good Life and the Human Good*. Cambridge, England: Cambridge University Press, pp. 169-189.
- Kitcher, Philip (1999) "Essence and Perfection" in *Ethics* Vol. 110, No. 1 (Oct., 1999) pp. 59-83.
- Loeb, Don (1995) "Full-Information Theories of the Individual Good" in *Social Theory and Practice* Vol. 21, No. 1 (Spring, 1995) pp. 1-30.
- Mill, J.S. (1979) *Utilitarianism*. Indianapolis, IN: Hackett Publishing Company.

- Moore, G.E. (1993) *Principia Ethica*. Rev. ed., Cambridge, England: Cambridge University Press.
- Nozick, R. (1974) *Anarchy, State, and Utopia*. New York, NY: Basic Books.
- Nussbaum, Martha, and Amartya K. Sen (1993) *The Quality of Life*. Oxford, England: Clarendon Press of Oxford University Press.
- Overvold, Mark Carl (1980) "Self-Interest and the Concept of Self-Sacrifice" *Canadian Journal of Philosophy* Vol. 10, pg. 105-118.
- Overvold, Mark Carl (1982) "Self-Interest and Getting What You Want" in Harlan B. Miller, and William H. Williams (Eds.) *The Limits of Utilitarianism*. Minneapolis, MN: University of Minnesota Press, pg. 186-194.
- Parfit, D. (1984) *Reasons and Persons*. Oxford, England: Clarendon Press of Oxford University Press.
- Paul, Ellen, Jeffrey Paul, and Fred D. Miller (1992) *The Good Life and the Human Good*. Cambridge, England: Cambridge University Press.
- Paul, Ellen, Jeffrey Paul, and Fred D. Miller (1999) *Human Flourishing*. Cambridge, England: Cambridge University Press.
- Paul, Ellen, Jeffrey Paul, and Fred D. Miller (1997) *Self-Interest*. Cambridge, England: Cambridge University Press.
- Quinn, Warren (1993), "Putting Rationality in its Place" in Warren Quinn (Ed.) *Morality and Action*. Cambridge, England: Cambridge University Press, pg. 228-255 .
- Railton, Peter. (1986) "Moral Realism" in (2003) *Facts, Values, Norms: Essays towards a Morality of Consequence*. Cambridge, England: Cambridge University Press, pg. 3-42.
- Railton, Peter. (1986) "Facts and Values" in (2003) *Facts, Values, Norms: Essays towards a Morality of Consequence*. Cambridge, England: Cambridge University Press, pg. 43-68.
- Rassmusen, Douglas B. (1999) "Human Flourishing and the Appeal to Human Nature" in Ellen Paul, Jeffrey Paul, and Fred D. Miller (Eds.) *Human Flourishing*. Cambridge, England: Cambridge University Press, pg. 1-43.
- Rawls, J. (1972) *A Theory of Justice*. Oxford, England: Clarendon Press of Oxford University Press.
- Rosati, Connie (1995) "Persons, Perspectives, and Full Information Accounts of the Good" in *Ethics* Vol. 105, No. 2 (Jan., 1995), pg. 296-325.

- Rosati, Connie (1996) "Internalism and the Good for a Person" in *Ethics* Vol. 106, No. 2 (Jan., 1996), pp. 297-326.
- Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
- Sen, Amartya K. (1980-1981) "Plural Utility" in *Proceedings of Aristotelian Society* Vol. 81, pg.193-215
- Shelley, Mary Wollstonecraft. (1818) *Frankenstein; or, The Modern Prometheus*. New York, NY: Scholastic Book Services.
- Shope, Robert (1978. "The Conditional Fallacy in Contemporary Philosophy" *Journal of Philosophy* 75, pg. 397-413.
- Slote, Michael (1990) *Goods and Virtues*. Oxford, England: Oxford University Press.
- Sobel, David (1994) "Full-Information Accounts of Well-Being" in *Ethics* Vol. 104, No. 4 (Jul. 1994), pg. 784-810.
- Sumner, L.W. (1996) *Welfare, Happiness, and Ethics*. Oxford, England: Clarendon Press of Oxford University Press.
- Velleman, D.J. (1990) "Is Motivation Internal to Value?" in his *The Possibility of Practical Reason*. Oxford, England: Oxford University Press.
- Velleman, D.J. (1991) "Well-Being and Time" *Pacific Philosophical Quarterly* 72, pg. 48-77.
- Wilde, Oscar, (1890) *The Picture of Dorian Gray*, in Richard Aldington (Ed.) *The Portable Oscar Wilde* (1946), New York, NY: The Viking Press, pg. 138-391.
- Williams, Bernard (1981) "Persons, Character and Morality" in Steven Cahn and Joram Haber (Ed.) *20th Century Ethical Theory* (1995), Englewood Cliffs, NJ: Prentice Hall, pg. 634-666.

BIOGRAPHICAL SKETCH

Anton Tupa received his bachelor's degree in philosophy with honors from the University of Kansas in May of 1997. He went on to get a master's degree in philosophy from the University of Kansas in May of 1999. He then studied applied ethics at Bowling Green State University for roughly four years. He transferred to the University of Florida in August of 2003 to complete work on his dissertation. He received his PhD in philosophy in May of 2006.