Three Kinds of Goodness for a Person

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Owen Christopher King, M.A.

Graduate Program in Philosophy

The Ohio State University

2016

Dissertation Committee:

Justin D’Arms, Advisor

Donald C. Hubin

Robert Kraut

Sigrún Svavarsdóttir
Abstract

I argue that three types of goods for a person—the goodness of her life, her well-being at a time, and the objects that are desirable for her—do not stand in the straightforward, intuitive relations philosophers have commonly assumed. I begin by developing an account of what I call value contribution principles. A value contribution principle says roughly that an increase in some property, all else equal, implies an increase in some value. Value contribution principles express relatively weak but nonetheless important evaluative claims. With near unanimity, philosophers have assumed, in essence, that a value contribution principle holds between a person’s well-being at a time and the goodness of her life, i.e., that an increase in her well-being during some period, all else equal, yields a better life for her on the whole. I argue that this is incorrect. Similarly, I argue that it is not the case that greater realization of what is desirable for a person, all else equal, increases her well-being. The result is that our notion of what is good for a person is less united and more equivocal than commonly recognized.
Acknowledgments

I wish to thank my advisor, Justin D’Arms, for extensive advice and feedback on all parts of this project over the course of many years. I also wish to thank my committee members, Donald Hubin, Robert Kraut, and Sigrún Svavarðóttir, for suggestions, criticism, and personal guidance. In addition, I am grateful for extensive written comments from Dale Dorsey on earlier drafts of Chapter 2 and Chapter 3. Chapter 3 has benefited also from feedback from Chris Pines and two anonymous reviewers for the journal *Ethics*. 
Vita

2001.........................B.S. Computer Science, University of Georgia

2002.........................A.B. Philosophy, University of Georgia

2006.........................M.A. Philosophy, University of Georgia

2006 to 2007...............University Fellow, Department of Philosophy, The Ohio State University

2007 to 2014...............Graduate Teaching Associate, Department of Philosophy, The Ohio State University

2014 to present............Visiting Instructor, Department of Philosophy, Oberlin College

Fields of Study

Major Field:  Philosophy
# Table of Contents

Abstract...............................................................................................................................ii
Acknowledgments...........................................................................................................iii
Vita.......................................................................................................................................iv
Introduction.....................................................................................................................1

Chapter 1: Value Contribution Principles..............................................................................6
  I. The idea of a value contribution principle.................................................................6
  II. The role and significance of value contribution principles......................................11
  III. What is equal?...........................................................................................................15
  IV. What may differ?.........................................................................................................21

Chapter 2: Well-being and Life Goodness...........................................................................32
  I. A value contribution principle relating well-being and life goodness.......................32
  II. Precedents for rejecting WL-CP?..............................................................................35
  III. Counterexamples to WL-CP.....................................................................................42
  IV. Key characteristics of well-being and life goodness.................................................48
  V. Well-being in a good life, reconsidered.......................................................................56

Chapter 3: The Untenability of Future-sensitive Synchronic Well-being.........................62
  Preface..............................................................................................................................62
I. Synchronic and diachronic well-being...............................................................65
II. The distortion problem..................................................................................69
III. The instability problem..............................................................................76
IV. Problems for past-sensitivity?....................................................................86

Chapter 4: Desirability as a Kind of Egocentric Value......................................91
Introduction..........................................................................................................91
Part I: Non-egocentric value.............................................................................95
I.1 Berlin’s conception of value.........................................................................95
I.2 A dispositional theory of value and imaginativist idealization....................96
I.3 Pluralism........................................................................................................103
I.4 Intersubjectivity..............................................................................................109
Part II: Desirability............................................................................................113
II.1 From the valuable to the desirable...............................................................113
II.2 Egocentric value: relational, not relative.....................................................118
II.3 Egocentric value need not be about well-being.........................................125
II.4 Parallels between desirability and (non-egocentric) value.........................130

References..........................................................................................................135
Introduction

At the most general level, this dissertation is about distinctions and relationships among kinds of value. The more specific focus is on what we can think of as kinds of egocentric value or personal value, that which is, in some sense, good for a particular person. I will argue that the goodness of a person’s life, the person’s well-being at a time, and the things that are desirable for the person, are three distinct kinds of egocentric value. And I will argue that these three kinds of value do not stand in the straightforward, intuitive relationships people have commonly assumed.

Before summarizing the project, it is worth looking a bit more closely at this notion of egocentric value. Compare these two evaluative statements: “The Sistine Chapel is beautiful,” and “Eight hours of sleep is good for Michael.” The statement about the Sistine Chapel is about non-egocentric value. According to that statement, the kind of value realized is not something to which any of us stands in any special relationship. The statement about the right amount of sleep for Michael is about egocentric value. The value involved is especially for Michael. Michael and I stand in the same relationship to the value mentioned in the first statement. Our relationships to the value in the second statement are quite different. Michael getting lots of sleep is significant for Michael in a way that it is not significant for me.
In general, we are in the realm of egocentric value when we are concerned with what is valuable for someone’s sake or when we are concerned with what is in someone’s interest. Consider my ill friend Horatio who, due to complications of diabetes, just had his foot amputated. An activity that has been a great source of joy throughout Horatio’s life has been the determined and careful cultivation of unusual and delicious varieties of peppers and tomatoes in his small, urban garden plot. For Horatio’s sake, I want him to be able to continue gardening, despite the fact that his mobility is now much more limited. Now, notice that continuing gardening is something we might consider good for Horatio in three different ways. First, Horatio is disposed to value gardening; it appeals to him. Second, Horatio will be faring better during the coming days and years if he keeps gardening. Third, it will be a better life for Horatio if he is not forced to cut short the pursuit to which he has been so committed for so long. These three perspectives on how gardening is good for Horatio correspond to the three types of egocentric value I will discuss: desirability, well-being, and the goodness of a person’s life.

The goal of Chapter 1 is to develop a theoretical apparatus (one which, as it happens, is not specific to egocentric value) that I put to use in the rest of the dissertation. This involves giving an account of what I call value contribution principles. A value contribution principle says, roughly, that an increase in some property, all else equal, implies an increase in some value. Value contribution principles express relatively weak but immensely important truths within the evaluative domain. It is ultimately because of the “all else equal” or “ceteris paribus” condition in value contribution principles that these principles can distill the general relationship between a property and a value to which the
property contributes. So, I take pains to provide a detailed account of exactly what a *ceteris paribus* condition allows and what it excludes.

One reason value contribution principles are useful is that they provide a way to make a distinctive sort of negative claim. To demonstrate that two values are *not* related by a value contribution principle is to show in a precise way an important sense in which the values are different. Intuitively, if two values are not related by a value contribution principle, it means that our respective concerns with the two values sometimes diverge. We see an important and surprising instance of this in Chapter 2. That chapter examines a particular value contribution principle relating two kinds of egocentric value, well-being and the goodness of a life. It is commonly thought that well-being is simply what makes a life good for the person who lives it. Even philosophers who have held that the relationship between well-being and life goodness is more complicated than this still typically assume that an increase in a person’s well-being during some period, all else equal, improves her life. In other words, it is common ground that there is a true value contribution principle that relates a person’s well-being at a time to the goodness of her life on the whole. But I argue that this is not the case. I describe and analyze examples that show that a person’s well-being can be increased without her life being thereby improved. The result is that the relationship between well-being and life goodness cannot be as tight as most have thought.

A crucial premise in the argument presented in Chapter 2 is that a person’s well-being at a particular time cannot be affected by events that take place after that time. In other words, the premise is that well-being is *not* *future-sensitive*. I offer a cursory defense of that premise in the course of presenting the arguments of Chapter 2. However, there is
more to be said about the issue. Chapter 3 offers an extended defense of the claim that well-being is not future-sensitive. The chapter takes the form of a critical discussion of a recent article by Dale Dorsey that advances a view according to which well-being is indeed future-sensitive.

Chapter 2 was about how a person is faring at a moment (i.e., her well-being) and the goodness of a person’s life. In Chapter 4, I distinguish another kind of egocentric value for a person. Intuitively, some things are desirable for a person, like friendship or good food, while other things are undesirable, like violence or receiving an insult. It may be tempting to think that the desirable things for a person are just those that, all else equal, increase her well-being. In other words, maybe we could define desirability by way of a value contribution principle relating it to well-being. In this chapter I give an account of desirability that makes it clear why this will not work. However, that temptation to think of desirability in terms of well-being is strong enough to prejudice a straightforward attempt to define desirability. Hence, I take an indirect approach. Drawing on work by Isaiah Berlin and David Lewis, I show a way to articulate a familiar from of non-egocentric value. Then I show how that account can be adapted to yield a definition of desirability. That puts us in a position to see why desirability cannot be understood as that which, \textit{ceteris paribus}, increases a person’s well-being or what, \textit{ceteris paribus}, improves her life.

In sum, we have three distinct kinds of goodness for a person—the goodness of her life, her well-being, and what is desirable for her—and my arguments show that they are not nearly as closely related as most people have thought. Hence, the notion of a person’s interest is more plural and more equivocal than commonly acknowledged. The upshot of
the arguments in this dissertation is that we are freed to theorize each of these values separately, without being constrained by spurious assumptions about how they are related to each other.
Chapter 1: Value Contribution Principles

I. The idea of a value contribution principle

Hedonistic utilitarians accept the claim that the moral goodness of a state of affairs depends on just the amount of pleasure realized in that state of affairs. But we do not have to accept this strong claim to think that these philosophers are, nonetheless, getting at something that is quite right: that more pleasure makes a state of affairs morally better. We can accept this weaker claim without being committed to thinking that pleasure is the only thing that affects the moral goodness of a state of affairs. Here are more careful formulations of the stronger and weaker claims, respectively:

\[(PM)\] If there is more pleasure realized in state of affairs A than in state of affairs B, then A is morally better than B.

\[(PM-CP)\] If there is more pleasure realized in state of affairs A than in state of affairs B and A and B are otherwise exactly similar, then A is morally better than B.

PM-CP is weaker than PM because PM-CP is qualified by a *ceteris paribus*, or all else equal, clause. So, a person who accepts PM-CP thinks that increased pleasure improves a state of affairs, from the moral standpoint, at least in those times where the increased pleasure comes at no additional cost. Now, though it is quite plausible, there is still room to disagree with PM-CP. But there is much less to disagree with, at least insofar as one
thinks that pleasure is morally valuable at all. In short, PM-CP expresses a claim that might be put this way: Pleasure contributes to the moral goodness of a state of affairs. Hence, PM-CP is an example of what I will call a value contribution principle.

We can identify plausible value contribution principles in other areas of ethical theory. As a principle of distributive justice, it is plausible that increased equality, if it is achieved by improving the situation of those worst off, makes a distribution of goods more just. Put more simply, equality achieved by benefiting the worst off increases justice. Or, more carefully:

(EJ-CP) If distribution A is exactly similar to distribution B, except that those worst off in B have in A a level of goods closer to the median level in B, then distribution A is more just than is B.

Like PM-CP, EJ-CP is not hard to accept. EJ-CP is entailed by strictly egalitarian views of distributive justice, and it is also entailed by Rawls’s Difference Principle. Thus, it is common ground between two otherwise competing views of justice. The views agree that, all else equal, elevating the position of the worst off to make them closer to the middle, is an improvement, from the standpoint of justice.

Even less controversial than either PM-CP or EJ-CP is a principle that says that a high level of well-being during a period in a person’s life contributes to the goodness of her life as a whole. As a contribution claim, it is not saying that well-being during discrete periods of the life is all that matters to the overall goodness of a life. It is compatible with this contribution principle that some other features besides the person’s levels of well-being at the different times affect the goodness of that life. But such a principle does say that, all
else equal, increasing a person’s well-being at some moment or during some period improves
her life on the whole. Put more carefully:

(WL-CP) If possible lives A and B are exactly similar, except for a period in which the
person has a higher level of well-being in A than in B, then A is a better life
than B.

As I said, WL-CP is even less controversial than PM-CP or EJ-CP. In fact, I am
not aware of anyone who has explicitly denied it. One of the main arguments of this
dissertation is that WL-CP is not true after all. But that will come later. For now, the
important point is to see how claims that a property or value contributes to some other sort
of value are aptly expressed through the sort of comparisons exemplified by PM-CP, EJ-CP,
and WL-CP, principles which use a ceteris paribus clause to express comparisons that isolate
the value contribution of a particular element. Along these lines, the general goal of this
chapter is to provide a thorough exposition of value contribution principles (VCPs). In the
remainder of this section, I will distinguish VCPs from similar and related principles. Then,
in Section II, I will explain the distinctive practical role and significance of VCPs. Thus,
these first two sections provide an introduction to value contribution principles. After that,
Sections III and IV get into the nitty-gritty. The ultimate purpose of those two sections is
to prove that the way I invoke VCPs in the rest of this dissertation, especially Chapter 2, is
consistent with the practical role and significance that I am generally attributing to VCPs.
The details in those sections show us how to figure out what comparisons are and are not
within the scope of particular VCPs.
Now, to see what is distinctive about VCPs, we have to recognize the significance of the *ceteris paribus* clause. To do so, it will be helpful to consider a similar sort of comparative evaluation that falls short of being a VCP. Suppose I went on a picnic today when the sky was overcast. And suppose I say, “It would have been a better picnic if it had been sunnier out.” Here, what I am claiming is that more sun would have contributed to the goodness of today’s picnic. However, I am not asserting some general principle about sunniness and the goodness of picnics. And that is good, because it seems quite clear that there is no true contribution principle in the vicinity. It is not always the case that, all else equal, sunnier weather improves a picnic. If there is already plenty of sunlight, any more might be too much. However, it might very well be true that, under normal circumstances, more sunlight will make a picnic better. This points us to an important caveat about *ceteris paribus* clauses. The term ‘*ceteris paribus*’ is commonly understood in two different ways. Sometimes ‘*ceteris paribus*’ is used to indicate that all other factors are constant across the objects being compared, i.e., *all else equal*. This is the more literal interpretation, and it is the one I intend here. However, sometimes the term is used to indicate a restriction of the scope of claim to just ordinary or normal cases. In a generalization which uses ‘*ceteris paribus*’ in this other way, the result is a rough generalization, one that holds for the most part, just not when abnormal factors are in play.¹

¹ Of course, if the normal and abnormal factors can be delineated, then we may regain generality by specifying these in the statement of the principle. But to do so would make the *ceteris paribus* (understood as ‘under normal conditions’) qualification unnecessary. For elucidating discussions of interpretations of *ceteris paribus* clauses, especially with regard to the formulation of scientific laws, see Schurz (2002) and Schurz (2014).
Value contribution principles are not rough generalizations. This distinguishes them from principles restricted to normal cases and also from another weaker (and less interesting) sort of principle—what we can think of as value correlation principles. A value correlation principle simply posits a positive statistical correlation between the increase in some property and an increase in some value. It seems quite likely to me that such a principle holds between increased sunshine and improved picnicking. However, establishing this would require empirical investigation.

Besides value contribution principles, there are other sorts of general connections between a property and a value which, nonetheless, may make it felicitous to say that the property makes the value or that an increase in the property is a part of the value. Here I have in mind what we can call ingredient relationships. Consider the relationship between sugar and the goodness of a cake. Sugar is an essential ingredient in a good cake. A cake with no sugar (if it even counts as a cake at all) is not a good one. However, it is not always the case that increasing the amount of sugar—even ceteris paribus—makes a cake better. Hence, we may have an ingredient relationship that holds generally between some property and some value without that property and that value being related by a true value contribution principle.

2 Value correlation principles may also involve ceteris paribus clauses. In an attempt to establish a statistical regularity, one typically takes pains to screen off the influence of confounding factors. Value correlation principles qualified by ceteris paribus clauses express relationships that do not depend on factors other than those cited by the principle.

3 An ingredient relationship does not entail a value contribution principle. But note also that the truth of a value contribution principle does not entail an ingredient relationship. And so we should not rule out the possibility that an increase in some property is guaranteed, all else equal, to increase some value, even though that sort of value may be realized in the total absence of that property.
To preview: In the next chapter I will argue that well-being and life goodness stand in an ingredient relationship, but not a contribution relationship. In general, this is the sort of configuration we are likely to find when the value in question is complex and is realized in an object that requires holistic evaluation.

II. The role and significance of value contribution principles

What do value contribution principles do for us? There are both epistemological and practical payoffs. The first epistemological point is straightforward: When a value contribution principle applies to a particular situation, recognizing this—whether explicitly or intuitively—increases our understanding of the value or values at hand. When we see that some particular factor contributed to the value of some object, we have a better sense of the object’s value. At the very least, recognizing the principle gives us a more fine-grained understanding of where the value came from, and it allows us to see at least one way in which the value would have differed had the circumstances been otherwise. Thus, appreciation of the counterfactuals expressed by contribution principles makes our evaluations richer and more nuanced.

Not only do value contribution principles enhance our appreciation of particular evaluations, they also structure our understanding of values in the abstract. An anecdote from a recent teaching experience is illuminating, I think. I was presenting an argument for the conclusion that unwillingness to provide certain sorts of justifications is bad. A central premise of the argument was that manipulation is bad. The undergraduate students, wary as they tend to be of any moral absolutes, jumped on this premise, offering counterexamples to show that the badness of manipulation depends on the situation, the interests of the
parties involved, etc. This reaction was, of course, on point. Few would accept the claim that all acts of manipulation are bad overall. I responded by saying that the premise should be understood as qualified by a *ceteris paribus* clause: If two actions are exactly similar, except that one is manipulative while the other is not, then the manipulative one is worse. Perhaps there are stronger true principles about the badness of manipulation, but this value contribution principle was all that was required for the argument to go through, and it proved unobjectionable to the students.

This is not an uncommon pattern in the search for general evaluative truths. First, we find we have an intuitive sense that there is some important connection between certain properties, some of them evaluative. But attempts to move from imprecise suggestions about a conceptual connection to some general principle are met with counterexamples and broad, apparently intractable disagreement. All too often, the knee-jerk response that follows is a rejection of any general connection (or, worse, skepticism about the existence of any general evaluative truths). Here value contribution principles come to the rescue. Though weaker than general principles that have no *ceteris paribus* qualifications, value contribution principles precisely describe general relationships between elements of the evaluative domain. For instance, consider the three fairly uncontroversial value contribution principles mentioned in the first section—principles relating pleasure and moral goodness, equality and justice, and well-being and the good life. This, to reiterate, is not to deny that stronger relationships may hold between some of these properties. But it is to insist that the relatively weak relationships expressed by true value contribution principles are substantive, general, and important. Unfortunately, we do not tend to notice these
principles as touchstones for evaluation. Likely, this is largely due to the fact that such principles are seldom formulated explicitly. Ironically, one reason they are seldom formulated explicitly is just that they are not usually sufficiently controversial to warrant thorough scrutiny. But these uncontroversial areas are exactly where we should look for robust, general truths.

In addition to these intellectual and theoretical benefits of value contribution principles, there are practical benefits as well. First, the principles, when true, provide a useful kind of guarantee. These principles warrant thoughts like this: If you’re going for \( Y \), then you can’t go wrong by just increasing \( X \). This sort of practical guarantee is a consequence of value contribution principles being general principles. When we have a true one, the consequent holds whenever the antecedent does. Increasing the property in the antecedent (given the ceteris paribus condition is met) guarantees increasing the value.\(^4\) This falls short of a recipe for optimization. A value contribution principle does not necessarily tell us the best way to boost some value, but it tells us one that is sure to work. It offers one way of making progress that cannot go wrong.

\(^4\) I admit that I am playing fast and loose here with an ambiguity in the word ‘increase’. The term can apply to either a temporal comparison, an increase between some time and a subsequent time, or to a counterfactual comparison, as a difference between two possible ways things could be. Clearly the latter is the sort of increase directly implicated in VCPs. However, from a practical standpoint, the temporal kind of increase is the one more straightforwardly realizable and, perhaps, more relevant. In general, the question of how we can exploit counterfactual comparisons of value in our practical deliberation is difficult and interesting—for some of the same reasons that questions about the application of idealized scientific laws formulated as counterfactuals are interesting. Although I have a bit more to say about the practical role of VCPs in the rest of this section, these points are largely beyond the present scope. My focus here is value, not practical reason. And, so, for immediate purposes, the important point is that sometimes we do, in fact, think in terms of VCPs and (somehow) make use of them in our practical reasoning.
Here are a few gems of folk wisdom that seem best interpreted as value contribution principles: “You can never be overdressed or overeducated.” “You can never have too many friends.” “You can never use too much butter.” These are, respectively, claims about how attire and education contribute to social or occupational success, how friendship contributes to goodness of a life, and how butter contributes to the tastiness of a dish. Of course, each seems subject to obvious counterexamples unless we understand these claims as limited by implicit *ceteris paribus* qualifications. (And, even then, these claims—especially the one about butter—are not beyond criticism.) But when qualified in this way, these sorts of principles offer a very useful form of practical guidance, showing us a risk-free way that we can be sure to further some end or realize some value.

This sort of guarantee can be particularly helpful in practical deliberation. If we find ourselves in a situation in which the value-contributing property can be increased *for free*, as it were, then, we should do it. Value contribution principles do not prescribe the pursuit of the contributing property at all costs; that is the point of the *ceteris paribus* qualification. But they do provide clear prescriptions when there is zero cost. Even if one has no higher goal than social ascent, she should not be improving her wardrobe at the expense of *everything* else. If one has poor communication skills or lacks funds for basic necessities, then she better not focus on her wardrobe at the expense of these things. But, on the other hand, if a person can become better dressed without holding back other improvements, then, according to our folk principle, she should do that.

This role for value contribution principles is not limited to those that are enshrined as popular sayings. There are plenty of value contribution principles we accept and appeal to
in decision-making. A common application is comparison shopping. Here is a mundane example. When shopping for a house, all else equal, the newer the roof, the better. Of course, the material and construction of a roof affect its performance, and some roofs are more attractive than others. But these factors are screened off the by the *ceteris paribus* condition. Another equally commonplace example: For an electronic device or piece of electrical equipment, the greater the battery capacity, the better. Unfortunately, larger battery capacity tends to come with increased size and weight, but not always. If size and weight and other factors are held fixed or can otherwise be factored out, then increased battery capacity is a clear benefit.

In general, then, these principles provide us with a way to compartmentalize some aspects of our practical deliberation. They allow us to make some choices about how to pursue Y, even in the absence of a complete plan regarding Y or even without all the information that would be required for a complete plan. They offer us a heuristic, and then some. Whatever else we will ultimately want to do to achieve Y, in the mean time we can worry about how to get X, as long as doing so does not close off other possibilities we might want to pursue later.

**III. What is equal?**

So far, I have offered a few examples of VCPs and have tried to show how they are important, but I have not yet precisely characterized the form of these principles. The difficulty in doing so is in saying exactly what the *ceteris paribus* qualification amounts to. We need to say what it means for one property to vary while “all else” is equal.
A VCP expresses a relationship between some property P and some value V. The general schema for a VCP is as follows:

\[
(CP) \quad \text{For any two objects A and B, if A and B are exactly similar except that A has a higher level of P than does B, then A has a higher level of V than does B.}
\]

At first glance, the CP schema may seem to characterize the content of VCPs precisely enough. What differences can there be between A and B? Well, just P, and everything else must be the same. But two simple observations show that the full story cannot be this simple. First, and most obviously, besides the difference in P, the CP schema must allow V to differ, too. After all, if exact similarity precluded a difference in V, then all VCPs (or at least the ones for which the antecedent was ever satisfied) would be false. So, V may vary. Exactly how V can vary and whether anything else is allowed to vary in that same way will be the topic of the next section. The second observation is that most properties do not vary totally independently of everything else, and so we may not be able to have a difference in P without other differences as well. That is the complication I will address now.

Suppose we have a situation in which A and B are exactly similar until the level of P in A is increased, and suppose the increased level of P in A causes changes in other properties that further distinguish A from B. Does the initial congruency of A and B suffice to make the CP schema applicable? The answer must be no. An affirmative answer would make any VCP subject to easily concocted counterexamples. To see this, consider the PM-CP principle introduced at the outset, which said that, all else equal, the existence of more pleasure makes a state of affairs morally better. Suppose that A and B are two exactly
similar worlds both of which happen to be such that the persons in the world become more malicious as they are more pleased. And now suppose that we adjust world A so that some of the persons in that world have more pleasure than before (and suppose we make no other changes). Then, when the A-world persons with more pleasure than their B-world counterparts behave more maliciously, world A may turn out to have a morally worse standing than B. But this is not the sort of case that should falsify PM-CP. The additional malice caused in world A—even though part of its cause was the extra pleasure in that world—is the sort of difference that a ceteris paribus clause is meant to exclude. VCPs are meant to capture the contribution of the contributing property per se, not other factors that contingently accompany it. In general, if the ceteris paribus clause failed to exclude comparisons where the objects are further distinguished by the causal consequences of the difference in the contributing property, then all VCPs would be subject to similar counterexamples.

So, at present, the conclusion to draw is that VCPs apply only to those cases in which the difference in P is causally isolated from anything that might produce further differences between A and B. Of course, this sort of causal isolation is unusual for most ordinary properties in most ordinary situations. And so this leaves only a narrow range of cases to which VCPs are strictly applicable, which some may find a troubling result. However, it is not a result that reduces the practical applicability of VCPs. I will return to this issue momentarily.

Even if we can entertain the possibility of an increased level of P independent of any other changes this might cause, increased P without any change in the properties on which
P supervenes is impossible. A human’s level of pleasure supervenes on her neural and chemical properties (and perhaps other things). So, in applying PM-CP, we cannot expect there to be a change in the person's pleasure without some change in these underlying properties. More generally, a VCP in which the contributor is a supervenient property must not exclude all comparisons in which there is some difference in the subvening properties. Otherwise the antecedent of the VCP would never be satisfied.

But how much difference in the subvening properties is allowed? This is a little complicated. We cannot allow just any adjustment of the subvening properties that suffices to realize the change in the contributing property. For a given change in the level of a contributing property that has a wide supervenience base, there may be many possible configurations of the subvening properties sufficient for realizing the change in the contributor. Some of these configurations might be such that they also include changes that are not required for the change in the contributor, and these unnecessary changes might still affect the value to which the contributor purportedly contributes. So what we should want, it seems, is the minimal change in the subvening properties that realizes the change in the contributor. But this is problematic, too. After all, many properties are multiply realizable. If a property can be changed by any of several adjustments to its supervenience base, then there may be no straightforward sense in which one of these adjustments is the minimal one. Probably what we should say is that a change resulting in a particular level of P satisfies the antecedent of CP if no proper part of that change would result in the same level of P, or something along these lines. Of course, this raises further questions about what it takes for one change to be part of another change, but these further
details can be safely ignored. For present purposes, as long as there is some coherent sense of *minimal change*, then this will be all we need to ensure determinate truth conditions for VCPs.\(^5\)

Consider now the issues about the practical applicability of VCPs which, as I have characterized them, have such strict antecedents. I have argued that we should understand the CP schema in such a way that the objects being compared do not differ with respect to the causal consequences of their difference in P, and such that they do not differ with respect to any lower level properties not strictly required to realize this difference in P. In general, the objects should have no dissimilarities that are merely contingently related to the way they differ with respect to P. But, in practice, we will seldom have reason—or even the opportunity—to compare objects that have this degree of similarity. This disparity between the strictness of VCPs and the messiness of ordinary circumstances does not limit the significance or even the usefulness of VCPs, though. A VCP expresses a particular sort of contributory relationship that holds between two properties, a relationship that holds even in situations to which the relevant VCP cannot be strictly applied. VCPs are counterfactuals; they say that *if* the antecedent held, the consequent would as well. The contributory relationship coincides with the truth of these counterfactuals. And so the epistemology of contribution is the epistemology of these counterfactuals. But once we have figured out that a contribution relation holds between properties, we need not fixate on the

---

\(^5\) Of course, if there are several coherent senses of *minimal change*, and these yield materially different truth conditions, then it will matter which of these we choose. I leave this as an open issue, with the hope that it is not one that ever proves particularly pressing.
technicalities of the relevant VCP. At that point we can operate with our intuitive understanding of the contribution relationship (as described in the preceding section).

In practice, figuring out whether a contributory relationship obtains typically involves settling for the examination of comparisons in which *most else*, as opposed to *all else*, is equal. Confirming or disconfirming a contributory relationship is a matter of figuring out whether it is P or something else that affects V. To do this we look for comparisons that approximate, as closely as possible, the strict *ceteris paribus* conditions just described, and then scrutinize the particular features of those comparisons that deviate from those strict conditions. This scrutiny aims to discover whether any of the factors that are only contingently related to the realization of P play any role in the explanation of the level of V. If what affects V is something contingent to the realization of P *per se*, then it is a case that neither confirms nor disconfirms the VCP. On the other hand, if no factors other than P explain the level of V, then the case is relevant to the confirmation of the VCP.

The point is that, for practical purposes, we do not really encounter (nor can we typically construct) comparisons in which everything required by the antecedent of a VCP is completely the same. And so we do the best we can, making sure that the contingent things that are not constant are not affecting the results. This is similar to how we deal with *ceteris paribus* generalizations in the sciences. For instance, the behavior of gases is such that, if the temperature of a gas increases and all else is equal, its pressure will increase. Schurz (2014: 1804) points to a similar *ceteris paribus* law relating temperature and volume.
were equal? Note that we could not even perform two observations for comparison without at least some difference in temporal or spatial location. So, strictly speaking, it is impossible to find a comparison in which only the temperature and the pressure vary; strictly speaking, it is impossible to find a comparison to which the *ceteris paribus* generalization actually applies. Of course, it is unlikely to occur to us that this would be any sort of problem because we are already assuming that time and location *per se* have no effect on pressure. But, on the other hand, if we perform the comparison with two side-by-side samples and either the composition of the gas or the volume of the container differs, then these differences may, as we know, affect the pressure. And, so, we would discount or ignore the results of such a comparison for purposes of confirming the *ceteris paribus* generalization.

In general, then, with both VCPs and scientific *ceteris paribus* generalizations, none of our actual comparisons fully meets the *ceteris paribus* condition. So, to figure out whether a given comparison manifests that relationship expressed by the generalization, we must rule out the possibility that any of the accidental differences between the things compared is affecting the property we are interested in. This sort of compromise between a principle's strict interpretation and the way we operationalize it is required for the deployment of abstract principles in the messy, actual world.

**IV. What may differ?**

I have argued that we should understand VCPs expressed in the CP form as applying strictly only to those comparisons in which the objects being compared do not differ with respect to anything bearing a merely contingent relationship to the difference in \( P \). Accidental differences that we must admit in actual comparisons are part of the practical
epistemology of VCPs, not in the scope of the principles themselves. Now, return to an observation from the beginning of the preceding section: Comparisons to which instances of the CP schema apply must allow at least one thing besides P to differ, viz., V. But this, of course, means something besides P is not the same. And this difference is definitely not just a feature of the messy circumstances of real-world application. Even applied to the most idealized comparisons, VCPs make no sense at all if V cannot change; the whole point is that V changes when P does. So, the question now is how to understand “all else equal” or “exactly similar” in a way that allows for a change in V.

In short, my claim will be that, in any comparison to which a particular VCP applies—any comparison that satisfies the VCP’s ceteris paribus clause—any difference in objects A and B (including, but not limited to, any difference in V) must be due to this: the difference in P together with the relations in which P stands to the objects’ other features which are constant across A and B.

What does it mean for other differences to be due to the difference in P? Well, we can answer this by saying what is the relationship we want to capture between the difference in P and the difference in V. Clearly, the relationship between the changes is not causal. As already emphasized, contingent consequences of a difference in P are precisely the sort of thing that makes a comparison run afoul of the ceteris paribus constraint. Instead, the relationship we want to allow is something like the grounding relationship, the relationship
sometimes signified by the phrase “in virtue of”. The difference in V must ultimately be
*grounded in* the difference in P.

My claim is that the *ceteris paribus* clause is consistent with not just the difference in V but also with any other differences that are similarly grounded. To fully state the view advanced in this section: We are trying to say to what comparisons a VCP—i.e., an instance of the CP schema—applies. For a VCP to be applicable to a particular comparison, the antecedent (which includes the *ceteris paribus* clause) of the embedded conditional must be true. My claim is that we should understand this antecedent in such a way that it being true is consistent with any difference between A and B (including, but not limited to, the difference in V) that is grounded in *the difference in P together with the background facts—the all else—that remain constant as P differs.* Or, for brevity, I will say any differences between A and B must be grounded in *the difference in P plus the background.* Or, to abbreviate more, any differences between A and B must be grounded in *the difference in P+.*

Again, according to the view I am advancing, not only does a *ceteris paribus* condition allow differences in V that are grounded in a difference in P+; it also allows similarly grounded differences in *other* properties. As a first step toward seeing why, note that allowing at least *some* such differences is unavoidable. It is impossible for differences in P not to ground differences in, at least, some relational properties involving P. After all, if P changes, *ipso facto,* we have a change in any property that involves P. For instance, if the

---

7 It has been widely argued that something like the grounding relationship between facts—the relationship that holds between facts when we say that one fact holds *in virtue of* another—is essential to much of philosophy. I intend my use of ‘ground’ to be consistent with that of Kit Fine, especially as in Fine (2012).

8 In light of the discussion of supervenience in the preceding section, here we should take the change in P to include the change in subvening properties required by the change in P.
level of pain increases and all else is equal, then, still, the ratio of pleasure to pain decreases. So, if VCPs did not allow differences in any properties besides the differences in V, there would be no comparisons at all to which they applied. And, so, for VCPs to make any sense, all else being equal must not rule out all other differences (differences besides the difference in V) grounded in the difference in P+.

But this does not yet show what I am ultimately claiming—viz., that the ceteris paribus condition is consistent with any and all differences grounded in the difference in P+. And there is some reason to hesitate short of that claim. Indeed, some of these additional differences might appear to be just the sort of interfering factors that ceteris paribus conditions were intended to rule out. In particular, we may wish to rule out differences which are grounded in the difference in P+ and which themselves ground changes in V. That is, we might wish to rule out cases in which changes in P+ indirectly ground changes in V by way of some mediating factor. We might think that mediating factors count as the sort of interfering factors that a ceteris paribus condition is supposed to bar.

Here, in the abstract, is the kind of case that concerns us. Suppose we have a particular comparison in which A has a higher level of P than does B, and suppose everything bearing a merely contingent relation to the difference in P is the same in A and B. Suppose also that the difference in P+ grounds a difference in some property M, making the level of M lower in A than in B. And suppose the lower level of M grounds a decreased level of V in A. (We might even suppose that there is a true VCP linking higher levels of M with higher levels of V.) So, M plays the role of mediating factor here. Now, finally, suppose that in all those other “ordinary” cases where the change in P does not change M
(and all else is equal), increased $P$ yields increased $V$. Then are $P$ and $V$ related by a true VCP? The answer is yes if and only if we have some compelling reason for saying that the difference in $M$ counts as a way in which all else fails to be equal, despite saying that a similarly grounded difference in $V$ does not count as a failure of equality.

To illustrate the issue about the admissibility of differences in mediating factors, let me introduce an instance of it that is central to the overall project of this dissertation. Recall WL-CP, the VCP relating well-being and the goodness of a person’s life. The principle says that if two lives $A$ and $B$ are exactly similar except that, during some period, the person’s level of well-being is higher in $A$ than in $B$, then life $A$ is a better life than $B$. According to our already settled strictures for exact similarity, in any case to which WL-CP applies, there is no difference between $A$ and $B$ that is merely contingently related to the difference in well-being. Now consider another possible sort of difference between $A$ and $B$. The level of well-being in $A$, in combination with features the two lives share, might ground the realization of other properties (mediating factors) in $A$ that further distinguish it from $B$. For instance, the extra well-being in $A$ might yield a life with a relatively incohesive narrative compared to $B$. This could be the case if the meaning of certain accomplishments common to the two lives depends on the person not faring very well during some parts of her life. Then the additional well-being could effectively reduce the life’s cohesiveness (i.e., add incohesiveness), which could reduce the overall goodness of the life. We can put aside, for the time being, worries about whether or not it is generally plausible that well-being levels can affect meaning and narrative cohesiveness in this way, and whether these features can affect the overall goodness of a life. The present question is
whether the extra incohesiveness in life A, on the supposition that it does play the role of a mediating factor which affects the goodness of that life, is consistent with the ceteris paribus condition in WL-CP. In other words, if we have a case in which a difference in well-being affects the goodness of a life due to the consequences of this well-being difference for the cohesion of the life, can this count as a test case for the contribution of well-being, all else equal, to life goodness?

My answer is yes. The question is about what sorts of comparisons are relevant to the truth of claims about value contribution. In other words, what comparisons of A and B are within the scope of a VCP? My answer is that the VCP applies to comparisons where, besides the difference in property P (the would-be contributing property), all else is equal. When is all else equal? I.e., when is the ceteris paribus condition satisfied? It is satisfied when any differences between A and B are grounded in differences in P+. In brief, the ceteris paribus condition is consistent with any differences grounded in differences in P+.

If someone wanted to disagree and hold that mediating factors of the sort described violate the ceteris paribus condition, what would be the rule by which the cases with mediating factors would be excluded? Part of the answer is that the rule would have to make reference to the role of constant background conditions, i.e., what distinguishes P+ from P. Ruling out cases where the difference in P grounds by itself a difference in some mediating factor M, which grounds by itself a difference in V, would definitely be ruling out too much. If we have a case in which the difference in P is the only factor responsible for a difference in M, which in turn is the only factor responsible for a difference in V, then it is very hard to see why that sort of case should not be within the scope of a VCP relating P
and V. To deny this would be to rule out transitivity of contribution as the explanation for the truth of a VCP. This would be a surprising and undesirable result. Intuitively, if X contributes to Y, and Y contributes to Z, then X contributes to Z. So, we should think that if a mediating factor ever makes a comparison outside the scope of a VCP, it will be due in part to the role of the constant background conditions in helping to ground the difference in the mediating factor.

Along these lines, note that in the kind of potential counterexample to WL-CP described above, it is not just the extra well-being that grounds the reduced cohesiveness. The other facts in the lives being compared, the facts that are constant between the two possible lives, are an essential part of the explanation. Following the earlier suggestion, maybe, for instance, something about a particular accomplishment (which is common to the two possible lives) is more meaningful if the person were faring worse at some particular point along the way. And, so, maybe extra well-being at that point would be a source of incohesiveness in that particular case. So, if we are inclined to think that some differences in mediating factors—like added incohesiveness—violate the ceteris paribus condition, then we should probably say it is not simply because it is a mediating factor, but also because the mediating factor is a difference due, in part, to the circumstances (the constant background conditions) of the particular comparison.

Generalizing this thought would yield the following proposal for excluding some mediating factors: A ceteris paribus condition is inconsistent with a difference in a feature M that affects V, iff that difference is grounded not just in the difference in P, but also partially in some feature of the constant background circumstances. More briefly, a mediating factor
is inadmissible if it is grounded in a difference in P+, not just the difference in P. As I have noted, even if we think that some comparisons involving mediating factors ought to be excluded from the scope of VCPs, there are many differences besides just the differences in P and V that a _ceteris paribus_ clause must allow if VCPs are to make any sense. The proposal here acknowledges this by narrowing the class of excluded comparisons to just those that have differences that, first, are grounded, in part, by some feature of the background circumstances shared by A and B, and, second, affect V. Though I think it is false, this proposal, I hold, is the most reasonable policy for interpreting the _ceteris paribus_ condition to exclude comparisons beyond those already excluded (according to the discussion in the preceding section) on account of differences in contingent factors.

I do not think this proposal is incoherent, but I think we should reject it on the grounds that it yields a much less useful and interesting conception of VCPs. That is because it excludes too many comparisons, narrowing the scope of a VCP far too much. If the proposal were correct, it would make almost any minimally plausible VCP impervious to counterexamples. To see this, observe how it underwrites a defense of what we should have thought would be a fairly ridiculous VCP. Consider this particular VCP relating the physical attractiveness of a man to how muscular he is:

(MA-CP) If A and B are two men who are exactly similar except that A is more muscular than B, then A is more physically attractive than B.

Suppose we wish to demonstrate the falsity of MA-CP with a counterexample. Begin by considering B who is a physically fit and fairly symmetrical fellow. Now consider A who is just like B except that the calf muscle on A’s right leg is twice as large as B’s right
calf muscle. Now, it might still be that A is highly attractive, but it is pretty plausible that B is more attractive. Man A is less attractive than B because A’s disproportionately large right calf looks a little silly. Or, less contentiously, A’s legs are asymmetrical in a way that is less attractive than if he had been more symmetrical. This, I would think, should suffice to show that MA-CP is false, as formulated. If we are interpreting *ceteris paribus* conditions in such a way that this is *not* actually a counterexample to MA-CP, then we must not be interpreting *ceteris paribus* conditions in the spirit of the sort of value contribution principles I have been describing in this chapter.

However, according to the proposal for excluding mediating factors, a defender of MA-CP can insist that our counterexample is inadmissible. She says that man A’s lower level of symmetry is a way in which it is not the case that all else is equal. The difference in symmetry is grounded in differences in the men’s calf muscles *along with* other bodily dimensions, which they share, and this difference in symmetry grounds a difference in the attractiveness of the two men. Hence, the comparison involves the sort of mediating factor that, according to the proposal, is inconsistent with the *ceteris paribus* condition.

If we were to accept this interpretation of the *ceteris paribus* condition and the ensuing defense of MA-CP, we would have to withdraw our purported counterexample. As I have already said, I think this is problematic and shows the implausibility of the proposed way of excluding from the scope of a VCP comparisons involving mediating factors. The proposal yields the wrong perspective on VCPs like MA-CP. Instead of understanding the VCP as ruling out of scope the comparisons involving mediating factors when these comparisons seem to contradict the VCP, we should accept that the VCP was subject to
counterexamples and view those comparisons that contradict it as providing guidance about how to formulate a more accurate VCP. Accordingly, we might replace MA-CP with a VCP that linked attractiveness to *symmetrical muscularity*, or, better, to *proportional, symmetrical muscularity*. If some such refinement proves impervious to counterexamples, then we have reached a true VCP. With respect to muscularity and attractiveness, I am not sure how far this would take us. I suspect that refinements would yield better and better principles subject to fewer counterexamples, but that no principle not invoking *attractive muscularity* (or something like that) will be flawless. And, of course, if the only true VCP we end up with says that an increase in *attractive muscularity*, all else equal, increases attractiveness, the alethic gain will be overshadowed by the loss in informativeness. But if that’s the best we can do, then it’s the best we can do. It may be that there is no informative, true VCP relating these two properties, and that in itself would be an interesting result.

In the next chapter, I will end up saying about WL-CP something similar to what I have just suggested about MA-CP. I will point to examples in which an increased level of well-being during a delimited period combines with other features of the person’s life that remain constant in such a way that it yields a life that is worse on the whole. And these are indeed genuine counterexamples—not comparisons ruled out of scope by the *ceteris paribus* condition. The upshot will be that we are forced to relinquish the general principle that increased well-being, all else equal, necessarily makes a life better.

---

9 This, of course, mirrors a dilemma for the formulation of scientific laws that apply only in normal conditions. If the conditions of normality can be specified, then we can dispense with the appeal to normality. If they cannot be non-circularly specified, then law is trivial. As noted earlier, the type of laws for which we sometimes face this dilemma are commonly called “*ceteris paribus* laws”, but then “*ceteris paribus*” is not being understood literally as “all else equal” but more like “all else being normal.” Again, see Schurz (2002) and (2014).
As for this chapter, in summary, we can say that a VCP applies to (and so potentially can be falsified by) a comparison between A and B if the *ceteris paribus* condition is satisfied. The condition is *not* satisfied by a comparison in which A and B differ in ways merely contingently related to differences in the would-be contributing property. However, the condition *is* satisfied if any differences between A and B are grounded in the difference in the would-be contributing property together with the features that are constant between A and B.
Chapter 2: Well-being and Life Goodness

I. A value contribution principle relating well-being and life goodness

Imagine Elliot, driving from his office to his wife’s soccer match. In one possible life, the drive is pleasant, he arrives for the kickoff, and he sees his wife slot home an early goal. Another possible life for Elliot is similar, except that traffic delays his arrival, and he does not see his wife score. We can imagine that, besides this difference, the lives are about the same. Elliot has a good relationship with his wife, and missing her score one goal would not affect that in any way. Still, for the period in which the two possible lives differ, Elliot has greater well-being in the life without the traffic delay. And furthermore, on the whole, the life in which he sees his wife’s goal seems like it should be slightly better for him than the otherwise similar life in which traffic delays him.

Examples like this lend credence to a claim about well-being that has assumed the status of a platitude: Well-being is what makes a life good for the person who lives it. But I

10 Assumptions along these lines are so pervasive that philosophers considering well-being and life goodness do not usually treat the two notions separately. Derek Parfit’s Appendix I to Reasons and Persons, which has been, since its publication, the standard starting point for discussions of well-being, introduces his topic by asking, “What would be best for someone, or would be most in this person’s interests, or would make this person’s life go, for him, as well as possible?” (1984, 493). More recently, at the beginning of Welfare, Happiness, and Ethica, Wayne Sumner states that welfare “attaches pre-eminently to the lives of individuals,” and he goes on to say “a person’s welfare is more or less the same as her well-being or interest” (1996, 1). In the introduction to Pleasure and the Good Life, Fred Feldman identifies “the Good Life” as “a life that is good in itself for the one who lives it,” and he notes, “Some philosophers speak here of ‘personal welfare’ or ‘well-being’. A good life, in this sense, would be a life that is outstanding in terms of welfare, or well-being” (2004, 9). Additional examples abound.
have become convinced that this is a serious mistake. It distorts the relationship between
well-being and the good life. To see this, a preliminary step is to recognize that the two
notions are not interchangeable. Just attending to the scope of the respective evaluations,
we can see one clear way in which well-being and life goodness differ. Well-being is a
transitory property of a person, a property she has (in some degree) at a particular time.
This is what allows a person to be faring better (have more well-being) today than on a
previous day. In contrast, the goodness of the person’s life pertains to the life as a whole.

Still, despite their distinct objects of evaluation and the rather obvious fact that they
are not completely interchangeable, it is natural to think that the relationship between well-
being and life goodness is very close. The relationship between improvements in each seems
especially tight. It is plausible that a person is benefited if and only if her well-being is
increased. Furthermore, it is plausible that a person is benefited if and only if her life is
made better than it would have been. So, we might be tempted to think that a person’s
well-being is increased if and only if her life is improved.

But I think this is incorrect. We have no guarantee that the facts that increase a
person’s well-being will result in a better life for her on the whole. Contrary to the
prevailing assumption, increased well-being over some stretch of time may not improve, and

---

11 Of course, this latter claim about benefit is plausible only if the life is better for the person whose life it is.
In general, when I talk about a life being good, better, or improved—even if, for brevity, I omit “for the
person”—the goodness I have in mind is goodness for a person, what is sometimes called “personal
goodness” (in contrast to impersonal values like moral goodness and aesthetic goodness). It may be
helpful to think of personal goodness along the lines of the account of welfare proposed by Stephen
Darwall (2002): as the sort of value it makes sense for you to promote for the sake of someone you care
about. (For a similar thought, see the account of personal value in Rønnow-Rasmussen 2011.) As I think
of it, personal goodness subsumes both what constitutes well-being and what makes a life good for the
person living it.
may even worsen, a person’s life on the whole, or so I will argue. This is so even if all other features of the life in question are held fixed. In other words, it is not the case that if a person’s well-being is increased, and all else is equal, her life will be improved. This, of course, is to deny the following value contribution principle I introduced in the preceding chapter:

(WL-CP) If possible lives A and B are exactly similar, except for a period in which the person has a higher level of well-being in A than in B, then A is a better life than B.

In the next section I show that even philosophers who have sought to problematize simple conceptions of the relationship between well-being and the good life do not go so far as to deny that additional well-being always boosts the overall quality of a life. Then, in Section III, I make my case that we should deny that increased well-being always improves a life. I describe and analyze examples that show that a person’s well-being can be increased without her life thereby being improved. Section IV extends the analysis of these examples by examining the key characteristics of well-being and life goodness that create space between these two sorts of evaluation. Ultimately, the two sorts of evaluation may diverge because well-being must be an essentially transitory property, while the goodness of a life depends on organizational features that may span a whole life. I conclude that, for any plausible theories of well-being and life goodness, the claim that increased well-being always improves a life will be susceptible to counterexamples. In the final section of this chapter, I offer suggestions about how we should think about well-being and the good life once the relationship between them has been loosened.
Theories of well-being are frequently offered with the promise of showing us what makes a life good for a person, but I think this is false advertising. Well-being and life goodness are distinct, though related, topics. But the problem is not simply that philosophers sometimes misrepresent their subject matter. More troubling is the recognition that intuitions about one kind of value are frequently invoked to support claims about another sort of value. The thought that well-being is what makes life good tends to guide and constrain—and thereby lead astray—our thinking about each evaluative category. We ought to correct this. If we cut the specious ties between well-being and life goodness, then clearer, more specific questions come into focus. The question, What makes life good? is replaced by three separate questions: What is well-being? What is a good life? and What is the relationship between the two? Once these three questions are no longer collapsed into a single question we will be in a position to rethink well-being while restricting any preoccupation with whole lives to its proper scope.

II. Precedents for rejecting WL-CP?

WL-CP is a fairly weak claim. It does not say anything about the relationship between the two evaluative categories, except that an increase in well-being, all else equal, entails an increase life goodness. It does not say anything specific about how different magnitudes of increase are related, let alone specify a function relating them. Nor does it offer any substantive explanation of why increased well-being should improve a life. Furthermore, the conditional points only in one direction; it does not entail the converse claim that life improvement requires increased well-being.
I am unaware of anyone who has presented an argument against WL-CP or any weaker claim about the relationship between well-being and life goodness. When philosophers have considered this relationship, they have tended to scrutinize stronger claims that sometimes have been accepted or assumed. For example, some have rejected some of the most straightforward explanations of WL-CP, but have done so without rejecting WL-CP itself. My present aim is to put WL-CP (and its rejection) into perspective by comparing it to other stronger claims about well-being and life goodness, which have been subject to greater scrutiny and criticism.

The most widely discussed claim about the relationship between well-being and life goodness is aggregationism. This is the claim that the goodness of a life is the aggregate well-being in the life. Take the entire span of a person’s life and partition it into short periods. For each of these short periods, multiply the level of well-being at that time by the duration of the period. According to the additivity principle that characterizes aggregationism, the sum of these products is the total goodness of the life.\(^1\)\(^\text{13}\)Aggregationism is perhaps the most literal way to make sense of the intuition that well-being is what makes life good. It says, more or less, that the goodness of a person’s life is

---

\(^{12}\) Michael Stocker is unusual in suggesting that WL-CP ought to be rejected. I take him to be rejecting WL-CP when he claims, “increases in the goodness of the constituent goods of a life need not make the life a better life” (1990, 323). I agree with him on this. Stocker does not, however, present an argument for the rejection of WL-CP. Instead, he assumes that it is false and invokes its falsity as an example to bolster his rejection of the general principle that increases in goodness of a part bring increases in goodness of the whole. He thinks the principle fails in those cases where the whole has its value in virtue of being an organic unity (1990, 301). One aim of the present chapter is to substantiate and elaborate these sorts of thoughts as they pertain specifically to well-being and life goodness.

\(^{13}\) A more cautious statement of the additivity principle would hold that the goodness of life is proportional to the sum of well-being, not simply equal to it. This preserves the core thought behind additivity, even if well-being and life goodness are distinct sorts of value.
constituted by (by being simply the aggregate of) the well-being in the life. However, though aggregationism and WL-CP each provide an interpretation of the intuition that well-being makes life good, we must avoid conflating the two claims. Aggregationism is a stronger claim than WL-CP; indeed, it straightforwardly entails WL-CP. Hence, we can reject aggregationism while leaving WL-CP unscathed. And this is just what we find in critical discussions of aggregationism.

Michael Slote (1983) famously has criticized utilitarians (e.g., Sidgwick) and contractarians (e.g., Rawls) alike for assuming the truth of aggregationism. Slote has argued that aggregationism is false on account of time preference, which is the claim that the value of some good in a person’s life is dependent, in part, on the point in the life at which it occurs. Two lives may have equal sums of well-being, even though the well-being is packed into the beginning of one and is distributed more heavily toward the end of the other. In the absence of other considerations, we may find we would prefer the life with more well-being toward the end. That is the idea behind time preference. If that is right, then aggregationism does not hold, and other factors besides the sum of well-being affect the value of a life. But time preference undermines aggregationism without posing any challenge at all to WL-CP. In fact, it makes sense for an advocate of time preference to be assuming WL-CP. We can understand time preference as the claim that increased well-being improves a life to a greater or lesser degree depending on when that well-being increase occurs. But it is natural to claim this while also claiming that, all else equal, increased well-being always improves (at least somewhat) a life.  

14 Perhaps according to an extreme version of the time preference, there are some periods in a person’s life for which the person’s well-being counts not at all in the overall goodness of the life. Maybe the first few
Whether or not Slote’s argument from time preference is convincing, there are other good reasons for rejecting aggregationism that do not amount to reasons for rejecting WL-CP. For instance, additive theories seem to be stuck with a problem analogous to Parfit’s Repugnant Conclusion: A sufficiently long life barely worth living will be superior to a shorter span of excellent living. One might attempt to improve on additivity by holding that the value of a life is, in large part, a function of the average level of well-being in the life. The resulting view would reject aggregationism. But, far from rejecting WL-CP, such a view would require it. After all, increased well-being during some period, other things being equal, increases average well-being in a life.

Views that equate the goodness of a life with aggregate or average well-being make the value of a life a function of magnitudes of well-being during the life. Since the values of the relevant function increase with any increase in well-being, the truth of WL-CP is ensured. Hence, we might think that denying that there is any function from well-being to goodness of a life would amount to a rejection of WL-CP. But this is not so. WL-CP is 

weeks or months during which a person is alive do not count. A view like this would actually contradict CP—since there could be increases in a person’s well-being without changes in the goodness of her whole life. This way of disputing CP seems to rely on a conception of life according to which there are periods which have no bearing on the quality of a life but are nonetheless genuine parts of the life. Whether or not this is plausible, it is not the sort of possibility that primarily interests me. I am arguing that there can be periods of genuine importance in a person’s life during which her well-being is increased but which reduce the quality of her life on the whole.

15 Feldman (2004, ch. 6) shows how to construct a hedonistic theory of well-being and the good life that preserves additivity while accommodating time preference.
17 I should point out that views that assess the goodness of a life primarily in terms of average well-being are subject to broadly Parfitian criticisms as well. A view that says the goodness of a life is just the average level of well-being in a life will say that a very short life with an extremely high level of well-being will be better than the same life extended by a period of great (though not extreme) well-being. This point parallels the point Parfit makes with the example of the world that ends with Adam and Eve (1984, 401-402).
compatible with—and may even be required by—views that deny that there is any such function. According to compositionalism, a view advanced by Johan Brännmark (2001), an individual’s well-being over some stretch of time depends on how that episode fits into the person’s life as a whole. Whereas aggregationist views start with an independent valuation of the parts of a life and determine the value of the life on that basis, a compositionalist view starts with the value of the life as a whole and evaluates parts of the life with respect to their role in the whole. Despite this difference, compositionalism is consistent with WL-CP and indeed seems to guarantee its truth. Suppose some particular facts increase a particular person’s well-being, in the counterfactual sense. In other words, if things had been otherwise, the person would have been less well-off during the time span in question. How are we to understand this increase, according to compositionalism? Well, according to compositionalism, a person’s well-being during some period of time depends on how the facts at that time fit into the rest of the person’s life. So, for facts to constitute an increase

18 The term “compositionalism” may be confusing. It is not supposed to suggest that the goodness of a life is simply composed from the goodness of the parts. On the contrary, this view proposes the opposite order of priority: The good of the parts is to be understood in terms of the good of the whole. Brännmark uses the term “compositionalism” because of what this view says about evaluation of lives. Compositionalism says that in the evaluation of lives, the composition (i.e., the arrangement) of the constituents matters; it is not just the independent values of the individual constituents that matter.

19 Stocker, citing precedents in Plato and Aristotle, sketches and seems to endorse a view of this sort: “To assess the effect of adding a good to a life, we cannot argue that since a good will be added, the life will be better. To see whether adding it is good—and whether more is better—we must evaluate the life that includes it. The goodness of adding the part thus depends on the goodness of the resulting whole” (1990, 300). It is not totally clear how this relates to Stocker’s apparent rejection of WL-CP (see note 12).

20 Brännmark does not give an account of how the goodness of some period in a person’s life is determined from the good of the whole. In theory, the decomposition that determines the values of the parts of a life could be such that the relationship between the values of the parts and whole life preserved additivity. However, the view would still be very different in spirit from typical additive views. Typical additive views tell us how to get the value of the whole from the independently determined values of the parts. Brännmark holds that no independent valuation of the parts is possible.
in well-being, those facts must fit into a life that is better than the otherwise similar life where those facts did not obtain. Hence, according to a straightforward construal of compositionalism, facts increase well-being only if the person’s life as a whole is better than it would have been had the facts been otherwise. This is just WL-CP. Perhaps some varieties of compositionalism will have a more complicated story about how the values of the parts of a life are determined from the value of the whole, but it is hard to see how a plausibly developed compositionalist view would not entail WL-CP in the way just described.

A commonality between compositionalist views and views like aggregationism is that these views understand well-being and the goodness of a life in such a way that one is derivative of the other. (The views differ primarily with respect to the direction of the derivation.) But, as we shall now see, denying that either is derivative of the other—i.e., accounting for each independently—is not tantamount to denying WL-CP. For instance, consider Shelly Kagan’s view about well-being and the good life. Kagan holds that the facts that determine a person’s well-being must be just the non-relational facts about that very person. To determine how good a person’s life is, though, additional facts may be relevant. For instance, on Kagan’s view, that I am having a pleasant experience is relevant to my well-being, but the fact that my pleasure is due to some false beliefs is not relevant, since the falsity of the beliefs is a relational matter. The falsity of the beliefs could, however, bear on the goodness of my life. Thus, Kagan explicitly rejects WL-CP’s converse (and therefore aggregationism), claiming, “certain changes might constitute changes in the quality of a person’s life without constituting changes in the person’s level of well-being” (Kagan 1994,
On this view, the well-being in a person’s life does not determine the goodness of her life. Furthermore, since the factors affecting well-being are specified without regard to their contribution to overall life goodness, Kagan’s view is not compositionalist.

Kagan’s view is an example of what we might call a superset view, since it holds that the factors that affect the goodness of a life include all those that affect well-being plus others. Arguably, this is the sort of view on offer also when David Velleman (2000) argues that the contribution of some period of well-being to the goodness of a person’s life is affected by how that episode fits in the person’s life story. These views point to features of good lives beyond just well-being. By holding that well-being and goodness of a life do not have the same supervenience base, superset views importantly weaken the connection between well-being and life goodness. In general, superset views allow a change in the overall goodness of a person’s life without any change in the person’s well-being, and, thus, contradict the converse of WL-CP. Still, superset views may be well-suited to preserve WL-CP itself. As long as everything that improves well-being also makes a life better overall, WL-CP holds. So, though each rejects the converse of WL-CP, neither Kagan nor Velleman disavows (or, for that matter, endorses) WL-CP.

---

21 See also Kagan (1992). Kagan’s restriction of the domain of facts relevant to well-being is contentious. We find opposition in both in supporters of additivity, like Feldman (2004, 109-114), and opponents of additivity, like Bigelow, Campbell, and Pargetter (1990, 127-134).

22 This thought will feature prominently in Chapter 3.

23 Like Kagan, Velleman rejects additivity and denies generally that the goodness of a life is a function of a person’s well-being. He also rejects the sort of compositionality offered by Bränmark. Though Velleman does not deny WL-CP, the sorts of cases and considerations he relies upon to undermine additivity would make it natural for him to do so. I return to this point at the end of section IV.

24 Though superset views need not pose a threat to WL-CP, they suggest another sort of view that would. If a superset view can make sense of how a life could be improved without any increase in well-being, then an analogous subset view (or, more generally, any view that posits factors that affect well-being but not the goodness of a life) could allow WL-CP to fail. According to a subset view, things could enhance well-
We have seen that superset views—like the other views canvassed in this section—need not threaten to WL-CP. I do think, though, that we should reject WL-CP. I think that WL-CP is false because of cases where certain facts make a person be faring better than she otherwise would have been, but those same facts make her life worse. This is a more pronounced divergence of well-being and life goodness than a superset view requires. Ultimately, something can be good for a person in one way and bad for her in another. In the next section, I offer counterexamples to WL-CP and examine their construction.

III. Counterexamples to WL-CP

In a genuine counterexample to WL-CP, we will have two lives that are almost exactly alike. They will differ during only a particular period, and the only difference during this period is that the person’s level of well-being is greater in one life than in the other. However, the life with the period of greater well-being will be worse overall. Of course, comparisons like this will be tough to come by. As I discussed in the preceding chapter, differences in certain properties will tend to come with other differences as well. So, for purposes of refuting WL-CP, what we want is a case in which there are few differences between the two lives, and none of the differences besides differences grounded in the levels of well-being affects the goodness of the two lives. The following scenario fits this bill, and thus serves as a counterexample to WL-CP.

being while having no effect at all on the goodness of a life. I do not know of anyone who has advocated a subset view. Likely, this is because most people assume that something like WL-CP is true. But if (as I claim) WL-CP is not true, then perhaps subset views deserve some investigation. At any rate, my argument against WL-CP does not require a subset view.

42
L1 and L2 are two possible lives for Frances. In both L1 and L2, Frances graduates from college with a bachelor's degree in fine arts during an economic recession. In both lives she is unable to support herself financially as an artist. In L1 but not L2, Frances's parents serve as her safety net, and she moves back home. At home she is comfortable and happy. For a year, she devotes herself to researching the weak economy’s effects on people, from factory workers to other artists. During that time she creates some excellent paintings that deal with the struggle of people in the poor economy, and, at the end of the year, she is recognized for her excellent artistic achievements. In L2, Frances does not rely on her parents as a safety net, and she is miserable for a year. She works two low-paying jobs. She is overworked, sleep-deprived, uncomfortable, and generally dissatisfied with her life. However, during that time she manages to produce some excellent paintings that deal with the struggle of people in the poor economy, and, at the end of the year, she is recognized for her excellent artistic achievements. In both L1 and L2, after gaining notoriety for her paintings, Frances becomes able to support herself as an artist, and she continues creating paintings with themes relating to poverty and duress. After the year in question, L1 and L2 are the same. Frances dies fairly young and never reflects on the conditions under which she created her first acclaimed paintings.

I am inclined to think that, for Frances, L2 is a better life than L1, even though during their only period of difference L1 includes more well-being than L2. This is not to say that the extra well-being in L1 is not valuable. It is valuable *qua* well-being, but not as an enhancement of Frances's life as a whole. L2 is better for Frances than L1 because of the way it is organized. A life in which success grows out of a period of hardship, a period of low well-being, has a structure that suits Frances particularly well.25 Moreover, we can imagine that if (contrary to the events of either possible life under consideration) Frances

25 Regarding the idea that what is good for a person fits her or is suitable for her in a certain way, see Rosati (2006).
herself had reflected on her reduced well-being in L2, she would have valued it as an element of her life. We can imagine Frances elevating her assessment of her entire life because of it, and I think we should agree that it is indeed a better life for her.

Now, admittedly, besides the different levels of well-being, there are other differences in these two possible lives for Frances. No theory of well-being will say that well-being is metaphysically basic; it supervenes on other properties. So, some of these subvening properties must differ between the two lives. Furthermore, it would be hard to make sense of a difference in well-being levels in the two lives if all the same events were common to both lives. Accordingly, during the year of difference, the two possible lives do not unfold in parallel. Hence, it is not the case, strictly speaking, that the two possible lives for Frances are exactly similar except for the difference in well-being. To determine whether the case bears on the truth of WL-CP, what matters is whether the comparative life evaluation is sensitive to these other differences. I said that a life in which Frances’s success grows from a period of low well-being has a structure that suits her. Since the subject matter of her paintings deals with certain kinds of hardship, her period of hardship during which she produced her first important work gives Frances’s L2 an agreeable sort of narrative unity. Given that the difference in narrative unity is grounded in the difference in Frances’s well-being (along with the other common features of the two possible lives), it is indeed specifically the well-being levels, not the other incidental differences, to which the comparative life evaluation is sensitive.26 Thus, for all practical purposes, besides the

26 Perhaps someone will be convinced that other differences between the two possible lives are also contributing to the narrative unity of L2. If that is right, the example should be revised to eliminate those differences. This would not affect the role of the example in the argument here. We would just be making the case more clearly within the scope of WL-CP, at the expense of making the comparison of the
difference in well-being levels, all else is equal. And so, this case serves as a counterexample to WL-CP.

Frances’s case is one in which a person, because of increased well-being, missed out on a better life. It would be a bit of a stretch to consider the increased well-being a *defect of* her life. It seems better just to say that, because of the increased well-being, her life on the whole was not as good as it could have been. We can, though, imagine other cases in which increased well-being would be most naturally construed as constituting a real demerit in the evaluation of the person’s entire life. That is what we have in the next example.

L1 and L2 are two possible lives for Valerie. In both L1 and L2, on a particular afternoon Valerie is driving on the highway where, off to the side, there is a motorist stranded with a flat tire. In L2, a butterfly splats on Valerie’s windshield and distracts her. She does not see the motorist, and she continues driving. In L1, there is no butterfly. Valerie sees the motorist, pulls over, and helps him change the flat tire. The motorist thanks Valerie, and this makes Valerie happy as she drives away. In both L1 and L2, Valerie then becomes captivated by a story on the radio. As a result, in neither life does she ever consider the motorist again. After that episode, L1 and L2 are the same. Several years later, in both lives, it comes to light that Valerie’s son has been molested by a neighbor. Unbeknown to Valerie and the neighbor, the neighbor is the motorist whom she aided years before.

It seems hard to deny that in L1 Valerie’s well-being is increased by helping the motorist. When she helped him, we may assume, the motorist was not yet the odious miscreant he later became. So, in aiding him, she achieved something worthwhile, and it was a gratifying experience. On the other hand, I think L1 is clearly worse for Valerie than two lives less striking.
L2. The abuse of her son is a great tragedy in Valerie's life, and her life is blighted by having helped the perpetrator. Now, in L1, after the day the motorist was stranded, she never thinks of him again. However, if in a variant of L1, she later thought about him and also became aware of his misdeeds, she would regret having had the benefit of the increased well-being that resulted from her interaction with him. I think such a response would be entirely natural and appropriate. And, hence, it seems that L1 is a worse life for Valerie than L2. Insofar as it is the difference in well-being between L1 and L2 that grounds this comparative evaluation of the goodness of lives, Valerie's case, like Frances's, serves as a counterexample to WL-CP.

In each of these two examples, I have tried to make the overall life evaluations plausible by suggesting that each protagonist would, if she reflected upon it, consider the increased well-being a demerit in the evaluation of her life as a whole. The idea is that a feature of a life is especially plausible as a characteristic of the good life in cases where the person reflects upon and values that aspect of her life. Indeed, it is not hard to imagine cases in which a person looks back on a time during which she was not as well-off as she could have been, and, in light of later events, feels grateful that things happened as they did. Similarly, a person might look back on something that constituted an improvement in her well-being, and feel regret or disappointment in light of later events. And the assessments implicit in these reflections are strong, if not decisive, evidence about how the past episodes would factor in overall evaluations of the lives. Furthermore, a person’s reflective endorsement buttresses the intuition that the value at stake is genuine goodness for the person, as opposed to some other sort of value that may be distinct, like beauty or moral
goodness. However, I doubt it is essential to counterexamples to WL-CP that the person in question, were she to reflect, would have any particular take on the relevant events. It is just that this sort of retrospection makes for compelling cases.

Because of the many different conceptions of the good life and well-being through which they might be interpreted, I expect my evaluations of the examples in this section to be controversial. For either example, and perhaps for both at the same time, it may be possible to identify a conception of the good life and craft a theory of well-being that make the example consistent with WL-CP after all. However, the question is not whether any particular example can serve as a counterexample to WL-CP in the context of all theories of well-being and life goodness. Rather, the question is whether every plausible combination of theories leaves WL-CP susceptible to some carefully constructed counterexample. I think that reflection on the structure of the examples in this section shows that the answer to this latter question is yes.

That said, I do think there are some conceptions of the good life and theories of well-being that do not leave WL-CP vulnerable to any counterexamples. But these conceptions are independently implausible. In the next section I will explain what

27 Regarding life goodness as goodness for the person whose life it is, see note 11.
28 Crucially, though, if retrospective assessment is to play a role in a counterexample to WL-CP, the retrospection must be merely counterfactual, not actual, within both of the possible lives compared. If the reflection actually occurred, this would be an additional difference between the two possible lives—which would mean the ceteris paribus condition was not satisfied. That is because if the person actually, at some later point, appreciated the time period in question, that appreciation would amount to an additional relevant difference between the two lives, and so it would not be the case that all else is equal. In a case that includes this retrospection, it would be open to a defender of WL-CP to argue that it was not increased well-being making the life worse but rather this other difference between the two lives. After all, whether and how a person actually reflects on her life may matter a great deal in the overall goodness of that life. So, it is crucial that the two lives being compared are no different in this respect. Hence, in the cases of Frances and Valerie, no retrospection or reflective assessment actually occurs.
characteristics of well-being and life goodness make WL-CP subject to counterexamples, and I will argue that we should not accept views that do not posit these characteristics.

IV. Key characteristics of well-being and life goodness

In the preceding section, I pointed out the significance of retrospection in the cases of Valerie and Frances. I said that if each had been able to reflect on her life she would have valued more the life without the period of increased well-being. This is important because (counterfactual) retrospective endorsement is a plausible indicator of a life’s value for a person. But even more important for present purposes is the broad temporal scope of such evaluations. Because they can take an entire life into view, retrospective evaluations are sensitive to the broad, relational features of lives. Counterexamples to WL-CP trade on these macro-level features; the counterexamples work because the goodness of a life depends on them.

If the right way of thinking about life goodness included no role for macro-level features of lives—focusing instead just on facts that affect a person’s well-being at various moments—then purported counterexamples to WL-CP (at least of the sort I have imagined) will be consistent with WL-CP after all. But I see little justification for this kind of restriction of the considerations bearing on life goodness. After all, among the properties of a life are the relations among the local features of the life. And assessment of a life, as with assessment of any complex entity, may depend on the relations among its parts. So it would be a mistake for anyone to assume, out of hand, that the macro-level, relational features were not relevant to the evaluation of a life. The default assumption should be that
they *are* relevant. A contrary claim would require considerable justification, which is not likely forthcoming.

We should suppose, then, that the right theory of the good life will take life goodness to depend, at least in part, on the broad, macro-level features of lives. But the conjunction of such a theory of life goodness with the claim that levels of well-being are *transitory* states of a person makes possible the disconnection I have described between increased well-being and life improvement. That is, this conjunction makes counterexamples to WL-CP possible. In light of this, an advocate of WL-CP might urge that we adopt a conception of well-being that makes it *not* merely a transitory state of a person. A more expansive theory of well-being might avert counterexamples to WL-CP.

And, after all, a person’s well-being at a time frequently seems to depend on facts that obtain at other times. For instance, the fact that I worked hard last month on a paper may—at least on some plausible views—be relevant to my well-being next month when I have the opportunity to present it. However, as I will argue now, the views of well-being that can avert counterexamples to WL-CP are implausible.

At this point, a few additional distinctions are required. First, we can distinguish theories of well-being with regard to *diachronic sensitivity*. If well-being is diachronically sensitive, a person’s well-being at a time may depend on facts that obtain at other times. According to some theories, well-being is not diachronically sensitive at all. A simple hedonistic theory of well-being is a paradigm case of a view according to which well-being is not diachronically sensitive.²⁹ Similarly, most other theories that satisfy the so-called

---

²⁹ That the correct theory of well-being is not diachronically sensitive is the claim Ben Bradley calls *internalism* about well-being at a time (2009, 18). This sort of internalism is a consequence of Bradley’s
Experience Requirement—the requirement that the things that can affect a person’s well-being are things the person experiences—are committed to a diachronically insensitive conception of well-being. In contrast, varieties of desire satisfactionism, provided they do not count only desires that are contemporaneous with their satisfaction, exemplify diachronic sensitivity.

We can further distinguish theories that admit diachronic sensitivity according to whether present well-being registers facts about the the past or about the future (or both). If well-being is past-sensitive, then it can be affected by events that have already taken place. And if it is future-sensitive, then it can be affected by events that happen subsequently. So, for instance, if well-being is past-sensitive, hard work in the past may affect a person’s present well-being when the work pays off. And, if it is future-sensitive, the recognition and praise that the work will later attract may affect her present well-being as she is doing the work.

These distinctions are important for present purposes because if well-being is future-sensitive, then examples like those I described in the previous section may pose no problem for WL-CP. In my examples, the protagonist’s well-being is increased at a particular time by facts that obtain at that time. Some subsequent events then stand in preferred variety of hedonism. But Bradley holds also that internalism is independently plausible, and he invokes it to criticize a broad class of diachronically sensitive views of well-being (2009, 21-28). In what follows, my case against diachronic sensitivity is restricted just to future sensitivity.

30 See Griffin (1986, 13-19).
31 Dale Dorsey (2013) takes up the question of what desire satisfaction theories should say about when a person with a satisfied desire is well-off. Dorsey defends the view that a satisfied desire makes a person well-off at the time she has the desire, even if the facts that satisfy the desire obtain at another time. Alexander Sarch (2013) defends a similar view. Dorsey (2015) relies upon and defends future-sensitive well-being in general. Chapter 3 is a sustained critique of Dorsey (2015).
32 Hence, on Dorsey’s view or Sarch’s view, well-being is future-sensitive (See note 31.)
evaluatively significant relationships to the increased well-being. These newer events relate to the earlier facts in such a way that the person’s life is worse than it would have been had there been no increase in well-being. So, for such counterexamples to work, the new events cannot alter the evaluative past, as it were, to make it the case that the original would-be well-being-increasing facts did not increase the person’s well-being after all. But an alteration of precisely this sort would be possible if well-being were future-sensitive.\(^{33}\)

According to that sort of view of well-being, it might be said that Frances really did not have less well-being during her year of poverty, precisely because the coherence of her life story would end up depending on the relation between how she was faring that year and her later career. And, similarly, according to such a view, perhaps Valerie’s well-being was not actually increased by helping the motorist, precisely because a future fact about that motorist stands in a troubling relation to future facts about Valerie’s life. Hence, the success of this type of counterexample depends on the claim that well-being is not future-sensitive.

But why should we deny that well-being is future sensitive? Certainly the fact that it may undermine counterexamples to WL-CP will not be a convincing reason to do so. On the contrary, since WL-CP is commonly taken for granted, the capacity to avoid counterexamples to it may seem an attractive feature. But these views have other, independent problems. The entirety of Chapter 3 is devoted to showing that a conception of well-being that makes it future-sensitive is untenable. There I lay out the problems that arise for attempts preserve aggregationism by way of future-sensitivity and the extreme (and implausible) complexity that comes from attempts to have well-being register the

---

\(^{33}\) Dorsey (2015) attempts to show how future-sensitivity could be invoked in this way to preserve aggregationism. I discuss this in detail in Chapter 3.
significance of temporally extended features of lives by way of future-sensitivity. At the present, I will simply raise a problem that offers intuitive reasons to think that future-sensitivity is not part of our ordinary concept of well-being. Though that is not yet a knock-down argument, I hope it will be somewhat persuasive nonetheless.

Roughly, the problem is that if a person’s well-being at the present depends on future facts, then her present well-being is not entirely present. And, so, in a situation where a person’s level of well-being changes over time, theories countenancing future-sensitivity reach equivocal conclusions about when that change occurs.

To see the problem, suppose a person had a baseline level of well-being for awhile, until her well-being increased at time \( t_1 \). And suppose it is the case (as future-sensitivity makes possible) that the well-being increase at \( t_1 \) was due to facts that obtained at \( t_1 \) together with facts that will hold at some subsequent time \( t_n \). Then we must say that although the person’s well-being changed at \( t_1 \), the change was not fully realized until \( t_n \). Not only is this odd on its face, it results in a revisionist interpretation of our ordinary judgments about a person’s well-being at a time. Imagine someone at \( t_2 \), where \( t_2 \) falls between \( t_1 \) and \( t_n \), wishes to assess and report the person’s change of well-being that occurred at \( t_1 \). But not all the facts bearing on the subject’s well-being at \( t_1 \) have yet been realized by \( t_2 \). So, when the assessor reports the subject’s well-being, how are we to understand this? Because the truth of her assessment depends on future events, it will be, at best, a projection or forecast of the person’s past well-being.\(^{34}\)

\(^{34}\) Note that this complaint is not specific to future-sensitive well-being. The problem (if indeed it is a problem) of forecasts of past states may arise with any future-sensitive property. As I mention below (note 35), many have considered the moral rightness of actions a future-sensitive property. I think the question of which properties (if any) are plausibly future-sensitive is an interesting and potentially important
Now, clearly any assessment of well-being will be prone to certain sorts of errors in observation and inference. But the predicament of the assessor at $t_2$ is worse than that. Note just how unusual the assessor’s projection must be. It is not an ordinary projection about the eventual result of an ongoing process; it is a future projection of the \textit{past} qualities of a person. The assessor might have to say something odd like this, “If events next month transpire as I expect, then she will have been faring well yesterday.” Whatever it is we are concerned about when we assess a person’s well-being at a moment, it is not something so sketchy. We think questions about a person’s present well-being are questions it makes sense to ask and answer in the present. I have never asked, “How are you today?” and received the response, “I should know by tomorrow,” and I would not expect to.\footnote{One might object that the situation here is no different than with act consequentialist theories of moral rightness, and so the problem here is not as bad as I am making it out to be. According to act consequentialism, the rightness of an action will depend on all the downstream consequences of that action, and so the rightness of the action at the time it took place is only as determinate as those future consequences. I do take this to be a something of a problem for act consequentialism, but I think the problem for future-sensitivity in well-being is worse. I assume actions are events, and it seems reasonable to evaluate an event in relation to other future events, especially those that it causes. It seems much more out-of-step with our ordinary thinking to think that a present quality of a person is dependent on future facts. The response, “I should know by tomorrow,” makes more sense in response to, “Did you do make the right choice?” than to “How are you today?”}

It is worth emphasizing that this problem does not depend on the details of any particular implementation of future-sensitivity. It affects the most extensively future-sensitive conceptions, like compositionalism, which say that well-being at a moment may depend on \textit{any} facts throughout the person’s life. It equally affects simpler views, like varieties of desire satisfactionism that say a person’s well-being at a time is a matter of whether her desires at that time will ever be satisfied. In general, it affects any view that question that deserves more sustained attention than I can offer here. For present purposes, it is enough if we recognize that future-sensitivity in well-being is counterintuitive.
countenances future-sensitivity precisely because any such view will, by definition, hold that it is possible for a person’s well-being at one time to depend on subsequent facts. But it is only this class of views that can ward off counterexamples to WL-CP, because only views of this sort can take into account, at the level of well-being, how subsequent events affect the evaluative significance of facts already realized.\textsuperscript{36}

I have argued that a plausible theory of the good life must be sensitive to the macro-level, relational features of lives. I have been urging also that we should not accept views that say that well-being is future-sensitive. These are the two basic claims on which my argument against WL-CP depends. Notably, these claims are \textit{no stronger} than the premises on which some cases against aggregationism have been built. In particular, very similar claims are essential to Velleman’s argument against aggregationism. Velleman explains, “Intuitively speaking, the reason why well-being isn’t additive is that how a person is faring at a particular moment is a temporally local matter, whereas the welfare of a period in his life depends on the global features of that period” (2000, 58). Though Velleman is distinguishing between well-being during short and long periods, and I locate the relevant

---

\textsuperscript{36} What about idealized desire satisfaction theories? These theories say, roughly, that a person’s well-being at a time depends on the satisfaction (at that time) of the desires she would have if she had a suitably ideal state of mind. (See Sidgwick 1907, 105-115, Railton 2003, and Sobel 1994.) Often, this ideal state of mind is taken to include “full” information. A person with full information is typically taken to have unlimited information about herself and her environment, especially about potential objects of desire and the means by which they may be had. Furthermore, idealized agents have predictive powers that far exceed those of ordinary persons. If so, then such theories may be able to avert counterexamples to WL-CP—if the agent’s ideal self desires only things which will improve the actual person’s life on the whole. So it may appear that these views can preserve WL-CP while avoiding the criticisms I have just described. The issues here are complicated, and a thorough treatment of them would require a paper. In brief, though, my opinion is this: At the point where fullness of information makes the agent’s desires sufficiently sensitive to goodness of a life to guarantee WL-CP, an idealized desire-satisfaction theory is subject to worries like those I expressed about future-sensitivity. In general, once a theory of well-being takes into account the entire domain of facts (including facts about the future) that could influence the evaluation of a whole life, well-being ceases to be an essentially transitory value.
distinction as between well-being and life goodness, the basic observations about the distinctive characteristics of the two kinds of evaluation are essentially the same. But Velleman applies these observations just to the sort of examples other opponents of aggregationism consider. He examines comparisons of possible lives in which, “the one life is better than the other even though they may include equal amounts of momentary well-being” (2000, 68). If instead we apply the observations to a different class of examples, as I have been urging, we reach the conclusion that we should reject WL-CP, not just aggregationism.

The difference between rejecting aggregationism and rejecting WL-CP is significant. Simply rejecting aggregationism permits us to continue understanding well-being and life goodness as two aspects of the same evaluative concern. But the rejection of WL-CP prohibits this, requiring us to recognize that well-being and life goodness are separable evaluations that do not necessarily vary in tandem. This separation of well-being and life goodness explains why aggregationism is false in the first place, and it ought to

---

37 Velleman is able to account for the distinction between well-being during a period and the goodness of a life as a whole, by taking the goodness of a life to be a special case of diachronic well-being, a person’s well-being over an extended period of time. And so our different ways of drawing the main evaluative distinction do not prevent substantive agreement on most points. That said, I do have a few reservations about the way Velleman draws his main distinction. For one thing, it seems to me that in order to make the distinction between synchronic well-being (well-being in a moment or short period) and diachronic well-being a sharp one, we would need a sharp distinction between short or momentary periods and longer, extended periods of time. That seems to me like a distinction that would be hard to draw in a principled way. Another option would be to think of synchronic well-being as a person’s well-being at an instant, and diachronic well-being her well-being over any non-instantaneous period. However, I am not confident that it makes sense to talk about how well a person is faring at an instant. And I worry that, if it does make sense, then well-being at an instant will be sensitive to the same factors as well-being during a short period. In contrast, a period in a life and a life as a whole seem to me to be more clearly distinct objects of evaluation (which are, nonetheless, frequently conflated).

38 The stronger result should be especially agreeable to someone like Velleman who is committed to there being two substantively different notions of a person’s self interest. (See Velleman 2000, 71.)

55
compel us to rethink the relationship between these two distinct evaluative categories. In the final section of this chapter, I will lay out several consequences and ramifications of rejecting WL-CP.

V. Well-being in a good life, reconsidered

I have been examining views about the relationship between well-being and the goodness of a life, and I have argued that WL-CP mischaracterizes this relationship. When we reflect on the possible ways of understanding this relationship, we find two general sorts of views that are compatible with WL-CP. One kind of view on which WL-CP comes out true privileges transitory goodness in a person’s life. These views start with a conception of well-being, saying what is good for a person during some limited timespan, and they understand goodness of a person’s life as a function of the person’s well-being at all the times throughout her life. The problem with these views is that they satisfy WL-CP at the expense of ignoring important global features of our lives. Accepting WL-CP along with a transitory conception of well-being yields an impoverished conception of life goodness. The other sort of view that accepts WL-CP takes the opposite tack: These views privilege goodness of the life as a whole and satisfy WL-CP by understanding well-being derivatively. These views understand a person’s well-being at a particular moment as whatever state would, in light of all the rest of her life, make the life a good one for her. By, in effect, defining well-being as whatever would result in a good life, these views satisfy WL-CP. But they do so at the expense of eliminating well-being as an independent evaluative category.

The views that privilege the transitory and the views that privilege whole lives share a common defect. They make one evaluative category a mere reflection of the other. Now
clearly there must be some close relationship between a person’s well-being at various times and the goodness of her whole life. After all, a life is composed of the moments in the life. What is good for the person during those moments ought to bear on what is good for her in the context of her whole life. But it seems to me that, in the absence of an abiding commitment to a principle like WL-CP, the reductionism manifest in both transitory-privileging views and life-privileging views is hard to motivate.

Once we give up WL-CP, how do we proceed? One predictable impulse is to adopt a mitigated variant of WL-CP. We could hold that to have well-being is simply to be in those states that *tend* to occur in good lives. Such a view would posit a positive statistical correlation between the things that increase well-being and the things that improve a life, and define well-being in terms of this relationship. This would be to claim that what I in Chapter 1 called a *value correlation principle* holds between well-being and life goodness. So, call this the *correlation view* of well-being. If we accept the correlation view, our investigation of well-being ought to uncover those general regularities that hold between the state of a person during a limited period of time and the goodness of her life as a whole. Such an investigation would be largely empirical, especially if we already have a credible way of evaluating life goodness.³⁹

---

³⁹ There are several ways to make the correlation view more precise. A fully developed view would specify the scope of the correlation between states of a person and goodness of possible lives. Are we looking for states correlated with good lives for any person? Or do we for each person come up with the set of states that are correlated with good possible lives for her? The former seems closer to the way people think about well-being in public policy contexts, and it is the view that offers the relatively straightforward empirical epistemology. The latter view comes with a difficult epistemology since it makes well-being depend on evaluations of possible complete lives for a particular individual.
At least one piece of the correlation view is almost definitely right. It would be shocking if it turned out that there were not an extremely strong positive correlation between increased well-being and life improvement. But the correlation view is not a satisfying theory of well-being, and I strongly doubt it expresses the full relationship between well-being and life goodness. The correlation view suffers from essentially the same defect that we find in those views that satisfy WL-CP by privileging the goodness of lives and relegating well-being to a derivative status. Like those views, the correlation view understands well-being fundamentally in terms of life goodness. The link is merely probabilistic, but, still, according to the correlation view, the significance of well-being is totally subordinate to that of life goodness.

The temptation to fall back from views that accept WL-CP to the correlation view is, I think, notable. Doubting WL-CP, we might cast around for another way to tether well-being to life goodness. I think this is the wrong response to the rejection of WL-CP. Our concern for how things are going for a person at the moment is not exhausted by how the moment fits into the larger whole of the person’s life. And vice versa: Our concern for the goodness of our lives is not exhausted by our concern for living well during all of life’s short periods.

I propose a different orientation. I do not think we are yet in a position to understand very precisely the relationship between well-being and goodness of a life. But I do think it is fair to postulate that, whatever else we learn about them, well-being is an essential ingredient in a good life. As I noted in Chapter 1, an ingredient relationship is an important way a property may be related to some value, even if no true VCP links the two.
Again, the analogy to the relationship between sugar and cake is instructive. Although sugar is an essential ingredient in cake, this fact does not give us the full story on the goodness of either sugar or cake. Sugar is good in a lot of ways, and, indeed, it can be good by itself. And, though a cake cannot be good without sugar, a cake that is simply a vehicle for sugar—though popular among children—is not an especially good cake.\footnote{The sugar and cake analogy is extensive, but it is not perfect. A low degree of well-being over long spans of a person’s life greatly diminishes the overall quality of that life. Similarly, a cake with little sugar is not as scrumptious as it could be. So far, so good. But recall the core (correct) thought of correlation views, that increases in well-being tend to be life improvements. In most cases an increase in well-being will, \textit{ceteris paribus}, improve a person’s life. Here the analogous culinary correlation is much weaker. Although many cakes will be improved by some additional sugar, many will not. It is much easier to have too much sugar in a cake than it is to have too much well-being in a life.}

The ingredient metaphor provides an alternative, though tempered, interpretation of the thought that well-being is what makes a life good. After all, there is an important sense in which sugar is what makes a cake good. But the ingredient metaphor does not involve the mistake of privileging either well-being or life-goodness in such a way that the other is to be understood derivatively. Instead, it posits related but distinct modes of evaluation. Of course, this does not yet tell us much about the two evaluative categories and the relationship between them. Most of the important questions await substantive accounts of one or both of the values. The ingredient metaphor provides, at most, a framework for such accounts. But this, in itself, is significant, because this new framework does not include WL-CP as a background assumption.

Theorizing without WL-CP in the background involves several challenges, but also some appealing prospects. Fundamentally, we have lost a constraint on our theories of well-being and life goodness. When accepting the assumption expressed by WL-CP, we may
allow our intuitions about well-being to be guided, in part, by our intuitions about what makes a good life for a person—that is, by our intuitive verdicts on comparisons of possible lives. So, without WL-CP, we lack what has been taken to be an important touchstone. I have argued, in effect, that we should replace WL-CP with two weaker constraints: that well-being is a transitory property and that it is an ingredient in good lives. But, for better or worse, I cannot see these replacements being as fertile as the assumption they replace.

Relatedly, without WL-CP, we are cut off from the most straightforward account of why well-being is important in the first place. The assumption expressed by WL-CP grounds a specious explanation of the significance of well-being. When we assume that well-being is what makes life good or what improves lives, since most of us think that the goodness of our lives is very important, we think we know why well-being is important. According to the standard assumption, well-being matters because it is important that lives be good for the people who live them. But the standard assumption is false, and so the simple explanation will not do.

Forced to relinquish a principle that has constrained our theorizing about well-being and also having lost a simple and intuitive (though mistaken) explanation of well-being’s importance, we can ask, without prejudice, what we really care about for persons during the particular moments of their lives. Our understanding of well-being should be guided by our answer to this question. No doubt we also care deeply—perhaps more than we care about well-being—about how a person’s life goes overall. But this concern ought not dominate *all* our thinking about what is good for a person. Perhaps our concern for a person in a moment pertains just to her pleasure and pain, or just to the satisfaction of her present
desires, or just to the realization of her considered values, or just to the proper functioning of her faculties, or to some combination of these. Whatever is the content of this concern, we should be able to reflect on it without being preoccupied by the organization of the rest of the person’s life. Thus unencumbered, our understanding of well-being can make legitimate progress. Likewise, we may enrich our understanding of what makes a person’s life good for her, once it is no longer a requirement that the account explain what is or is not a good state for a person at a particular time. I think that it will be only once we have made progress in these domains, understood separately, that we can give a more complete answer to the question, with which this chapter has been primarily concerned, about the relationship between well-being and the good life.
Chapter 3: The Untenability of Future-sensitive Synchronic Well-being

Preface

In the preceding chapter, I argued against WL-CP, a VCP relating a person’s well-being at a time and the goodness of her life on the whole. My argument relied crucially on the claim that well-being is not future-sensitive. That is, a person’s level of well-being at a particular time cannot depend on what happens subsequently. If well-being were future-sensitive in this way, a person’s well-being at a time could reflect the relationships that hold between facts at that time and later events. Hence, a person’s well-being at a time could reflect the relational factors that bear distinctively on the overall goodness of the person’s life. This would make it possible, in principle, for a theory of well-being to validate WL-CP.

In the same way that it might serve to save WL-CP, future-sensitivity might also save aggregationism. As I pointed out in the last chapter, the considerations on which I have based my rejection of WL-CP are the same sorts of considerations on the basis of which others, especially Velleman (2000), have rejected aggregationism. Hence, the strategy of having well-being register the significance of temporally extended features of a life via future-sensitivity might also be part of an attempt to save aggregationism. Indeed this is the strategy employed in Dale Dorsey’s 2015 article, “The Significance of a Life’s Shape.” In that article, Dorsey argues that certain observations that have been taken to threaten
aggregationism are compatible with it after all. Specifically, it is often thought that the sequence of good and bad times in a person’s life (the “shape” of her life) affects how good her life is on the whole. This seems to spell trouble for aggregationism simply because aggregation takes no account of temporal order. Dorsey argues that, on the contrary, it is possible to reconcile aggregationism with the claim that the shape of a life affects its value. The key to the reconciliation is the claim that well-being is future-sensitive. If well-being were future-sensitive, then a person’s well-being at a time could reflect how events at that time fit into the larger temporal sequence of life events. This would make it possible, in principle, to preserve aggregationism while accommodating the observation that the shape of a person’s life is evaluatively significant.

The present chapter takes the form of a critical discussion of Dorsey’s article. The main thrust of my criticism of Dorsey’s article is this: A conception of well-being that countenances future-sensitivity entails intractable complexities, to the detriment of our ability to evaluate what is in a person’s interest. Thus, my criticism of Dorsey also constitutes my case against future-sensitive well-being in general.

Rather than adapt the content of my critical discussion of Dorsey’s article, so that it picks up right where my last chapter left off, I have left the discussion in the form in which I expect it to be published independently of this dissertation. I do not think this should be the source of any significant confusion, and I see little downside to this way of presenting the material. Still, a few remarks are in order.

First, in this chapter, so that my terminology corresponds with Dorsey’s, I distinguish between a person’s synchronic well-being and her diachronic well-being. A person’s
synchronic well-being is the same as what I have been calling her *well-being at a time* or simply her *well-being*. In contrast, when we talk about a person’s diachronic well-being, we are talking about how she is faring over an extended period of time, perhaps over her entire life. So, it is natural, and unproblematic for present purposes, to understand the goodness of a person’s life, on which I was focused in the preceding chapter, as a special case of her diachronic well-being.

Could the argument of the preceding chapter have been expressed in terms of the relation between synchronic and diachronic well-being, instead of the relation between well-being and the goodness of a person’s life? Perhaps. However, I have chosen not to do this because I harbor some reservations about whether the distinction between diachronic and synchronic well-being can be maintained.\(^{41}\) Luckily, nothing in the present chapter depends on our ability to cleanly distinguish diachronic well-being. Furthermore, if we do assume that the goodness of a person’s life is the same as her diachronic well-being over the period coterminous with her life, a VCP linking synchronic well-being and diachronic well-being would be stronger than an otherwise similar VCP linking synchronic well-being and the goodness of a life. Hence, rejecting the latter VCP, as I have argued we should, is sufficient for rejecting the former as well.

Finally, in order for my critical discussion of Dorsey’s article to be self-contained, I have had to repeat my introduction of the notion of diachronically sensitive (synchronic) well-being, and the two more specific notions of past-sensitivity and future-sensitivity.

---

\(^{41}\) I say a bit more about the reasons for my reservations in note 37.
This was an unavoidable cost of preserving the critical discussion in its original form, but I do apologize to the reader for this bit of redundancy.

I. Synchronic and diachronic well-being

In an influential paper, David Velleman has drawn a distinction between synchronic and diachronic well-being. A person’s synchronic well-being is how well she is faring at a moment (at an instant or over a few minutes or, maybe, hours), and her diachronic well-being is how she fares over an extended period of time (over the course of days, years, or even a lifetime). It is plausible that a person’s level synchronic well-being and her level of diachronic well-being do not depend on all the same factors. A person’s synchronic well-being at a moment, one might think, depends only on what is the case at that moment—paradigmatically, mental representations and feelings realized at that moment. In contrast, diachronic well-being plausibly depends, in part, on relations among states of affairs at different times. For instance, it is commonly assumed that, all else equal, a long period of steadily increasing synchronic well-being is preferable, from the standpoint of diachronic well-being, to a symmetrical period of steadily decreasing synchronic well-being. Arguably, also relevant to diachronic well-being are various more complicated sorts of narrative features that may encompass disjoint periods in the life of an individual. For example, whether some events in a life fit into a triumphant story or a tragic one may affect the person’s diachronic well-being over the period spanning those events. Since diachronic well-being depends on factors that do not affect synchronic well-being, we have some reason

---

42 Velleman (2000).
43 Michael Slote (1983) is a famous proponent of this sort of claim. It is a consequence of pure time preference—the claim that the value of goods depends on when (relative to a life span) they take place.
to think these are genuinely distinct kinds of value. And, even if we hesitate to accept that there are two different kinds of value here, we still feel pressure to conclude that a person’s diachronic well-being over a period is not the same as her aggregate synchronic well-being during that period.\footnote{Besides Velleman (2000), prominent opponents of aggregationism include Slote (1983), Stocker (1990), Bigelow, Campbell, and Pargetter (1990), Kagan (1992), and Brännmark (2001).}

But there are several ways to resist these conclusions. One is simply to deny that any diachronic factors affect diachronic well-being. In other words, all the properties on which diachronic well-being depends are synchronically realizable. Then it is possible to hold that the same factors affect both synchronic and diachronic well-being, making it much easier to see how the latter could be an aggregate of the former.\footnote{This may not be as far-fetched as it might sound at first. Securing aggregationism by denying that narrative relations among events have any direct effect on diachronic (or synchronic) well-being does not rule out significant indirect effects. Rosati (2013) suggests that how we construe the narrative relations among events in our lives can profoundly impact how we experience these events and thereby affect our well-being.} Another way is to accept that diachronic relations do affect diachronic well-being, but claim that they also affect synchronic well-being. This alternative way of coordinating the two sorts of well-being is the strategy recently developed by Dale Dorsey, especially in his “The Significance of a Life’s Shape” (2015).\footnote{Dorsey (2013) provides an independent defense of some of the same theses—with special attention to desire-satisfaction views of synchronic well-being.} As Dorsey puts it, “the narrative relations borne by individual, temporally discrete events help determine not only the contribution of these events to diachronic welfare but also the contribution of these events to synchronic welfare” (2015, 325).\footnote{Like most philosophers writing about these subject, Dorsey uses ‘welfare’ and ‘well-being’ interchangeably. I too consider the two terms interchangeable, though I tend to use the term ‘well-being’.} Since, according to this sort of view, the diachronic factors affect synchronic, not
just diachronic, well-being, Dorsey argues that an aggregationist conception of the relationship between the two may be workable after all. But, of course, this picture requires a conception of synchronic well-being according to which present well-being is affected by events that take place at other times.

I will say that synchronic well-being is *diachronically sensitive* if a person’s synchronic well-being at a time may be affected by events at other times. There are two mutually compatible kinds of diachronic sensitivity: past-sensitivity and future-sensitivity. If synchronic well-being is *past-sensitive* then it can be affected by events that have already taken place. Compare two scenarios. In one, I have long yearned to visit a particular region, and then I travel there and enjoy my stay. In another, I have never had a strong opinion about that place, and then I travel there and equally enjoy my stay. If synchronic well-being is past-sensitive, it need not be true that the levels of my synchronic well-being during my trip would be the same in both scenarios. A desire-satisfaction view that accepted past-sensitivity could hold that my well-being is elevated in the first scenario precisely because of my prior desire to be doing what I am doing. Although it is not especially intuitive, we can also imagine how past-sensitivity might look in a hedonistic view of well-being. Perhaps the thought would be that the happiness that follows a period of sadness is somehow a greater benefit, hence yielding greater synchronic well-being, than equal happiness that was preceded by more happiness.

If synchronic well-being is *future-sensitive*, then how well I am faring in the present may depend on what happens at some point down the road. Suppose that I presently desire

---

48 Dorsey does not attempt to make a positive case for aggregationism. His goal is to establish that acknowledging the importance of life’s shape does not rule it out.
that a particular political candidate, for whom I have been volunteering, win an upcoming election. Consider two possible futures. In one, my candidate wins, takes office, and ushers in reform. In the other, she loses and we are stuck with the status quo. One might hold that I am faring better now if my actual future is the one in which my candidate wins. This is what we might say according to an alternative variety of desire satisfactionism, one that says that future satisfaction of my present desires makes me better off in the present.49 According to such a view, well-being is future-sensitive. Also, though I doubt there is much to recommend it, it is conceivable that future-sensitivity could be part of a view that ties synchronic well-being to happiness. Such a view might hold that a period of happiness that precedes a precipitous decline is worse in terms of synchronic well-being than a period of similar happiness that precedes more happiness.

I am not inclined to think it is problematic to hold that synchronic well-being is past-sensitive, and, though it is not my primary focus, I will have a bit more to say about that issue in Section IV below. Future-sensitivity is my primary concern. It is quite problematic to hold, as Dorsey does, that synchronic well-being is future-sensitive.50 In the next couple of sections I make a case against a future-sensitive conception of synchronic well-being. In Section II, I show how future-sensitivity robust enough to secure aggregationism requires a revisionist interpretation of the shape of life considerations that were a primary motivation for it in the first place. Section III shows how, according to the most natural ways of implementing it, future-sensitivity introduces an unanticipated and


50 See Brännmark (2001) and Sarch (2013) for other views according to which synchronic well-being is future-sensitive.
unacceptable sort of complexity into our evaluations of well-being—complexity of a sort that threatens even the determinacy of these evaluations.

I should emphasize one point about the case I will be making against future-sensitivity. To adopt a future-sensitive conception of synchronic well-being would reduce, or perhaps eliminate, any significant role for an independent diachronic conception of well-being. After all, future-sensitivity promises to account for diachronic considerations within the synchronic. One way of opposing this sort of picture is to demonstrate the theoretical benefits of taking synchronic and diachronic well-being as two separate evaluative categories. The thought is that countenancing two sorts of value allows us a nuanced and accurate view of self-interest and our judgments about it. I am quite sympathetic to this, but Velleman and others\(^{51}\) have already offered arguments along these lines. My argument here is negative: A conception of synchronic well-being sufficiently future-sensitive to reflect the value of significant diachronic features of lives is untenable.

II. The distortion problem

Imagine a case in which the shape of a person’s life—i.e., the sequence of rising and falling levels of synchronic well-being—seems evaluatively significant. First, consider a person whose life has a negative trajectory; suppose her level of synchronic well-being is high during youth, middling during her middle age, and low during her later years. Compare this to a person whose levels of well-being are the same but come in the reverse order, so that she fares poorly when she is young but her level of synchronic well-being follows a steady upward trajectory over the course of her life. And suppose that, to whatever

\(^{51}\) See, especially, Kagan (1994).
extent possible, the two lives are otherwise equal. Then, according to the common assumption mentioned at the outset, that increasing synchronic well-being is better than decreasing, the life with the upward slope is better.

If it is true that the upward-sloping life is better, how are we to account for this? Since the two lives are symmetrical, with the same levels of synchronic well-being in each, the difference between an upward slope and a downward slope is not reflected in aggregate synchronic well-being. Because the shape of a life depends on the relations among the states of the person at different times in her life, it is precisely the sort of feature that is not registered in synchronic well-being levels, at least insofar as synchronic well-being is not affected by diachronic features of lives. But if synchronic well-being is diachronically sensitive, as Dorsey argues it may be, then it can register the diachronic features. In particular, if synchronic well-being is future-sensitive, then it may be that the good times early in the downward-sloping life, on account of later misfortune, count for less than they otherwise would have. This would reduce the aggregate synchronic well-being of the downward-sloping life, effectively lowering its value on the whole. Similarly, the bad times early in the upward-sloping life count for more, increasing the aggregate synchronic well-being of this life and raising its value on the whole. And so aggregation may, after all, yield a result that matches our intuitive judgment that the upward-sloping life is better. Thus, it seems, with future-sensitivity we might preserve aggregationism.

But notice what just happened. The account of the case required what amounts to a redescription of the case itself. The task was to explain the divergence of our evaluations of two lives in which the levels of synchronic well-being are the same but come in reverse
temporal order. But then the crucial move was to deny that the levels of synchronic well-being are the same in the two lives. If diachronic factors are reflected in synchronic well-being in a way that adjusts the synchronic well-being levels to make the respective totals differ between the two lives, then the two lives were not symmetrical in the way required to generate the puzzle in the first place.

Now, one might think this is a desirable result: the puzzle has been dissolved. But it has not been dissolved in the right sort of way. Instead of accounting for the value of the shape of a life, Dorsey’s proposal has us reassess the shape itself. Consider the life we imagined sloping upward, and continue supposing that an upward slope is good. Because of future-sensitivity, the early times in the life realize more synchronic well-being than it first seemed they did. This makes those early times higher, which means the life may not slope upward after all; or, at least, the slope will be less steep. Thus, the very feature whose value we were trying to capture has changed.52 The life does not have the shape we thought it had. In short, the problem is this: If synchronic well-being is future-sensitive in the way required to reflect the value we attribute to the shape of a life, then a life’s shape may differ from the shape to which we had been attributing value. Since the issue is the way that future-sensitivity effectively bends the curve of a person’s well-being level over time, call this the distortion problem.

What if the value of the upward slope of the life adjusts all the synchronic well-being levels in the life by the same amount? Then, the life would still realize the same upward slope, just uniformly shifted higher. So, the valuable feature, it seems, would be unchanged. Dorsey suggests something much like this (2015, 236, note 36). But this would not solve the problem. What ultimately matters is the sequence of absolute, not relative, levels of synchronic well-being. To see this, note that a sequence that goes from very bad to pretty good does not have the same significance as a sequence that goes from pretty good to really great, even if the distance between very bad and pretty good is the same as the distance between pretty good and really great.

52
To see that the distortion problem is genuinely problematic, and especially so for Dorsey, it is helpful to see exactly how it emerges in the context of his article. Dorsey’s main task is to investigate what he formulates as the *Shape of a Life Hypothesis*.

*Shape of a Life Hypothesis* (SLH): The temporal sequence of good and bad times in a life can be a valuable feature of that life as a whole.\(^{53}\)

SLH is not a hypothesis in the standard sense of a claim that we aim to confirm or refute. For Dorsey, that we make judgments that align with SLH is a datum in need of explanation. We find widespread agreement about a class of evaluative judgments that are well-summarized by SLH. For instance, we tend to think that lives that improve as they go along are, all else equal, better than lives in which the person becomes worse and worse off. What explains this pattern of evaluation? It is a good question, and Dorsey aims to answer it. Furthermore, he hopes his answer will put us in a position to decide whether attributing value to diachronic features of lives could be consistent with aggregationism.

The central claim in Dorsey’s account of our SLH thinking is that the sequence of good and bad times is not valuable in itself, but is instead *signatorily* valuable; that is, the sequence is a reliable indicator or *sign* of something else that is intrinsically valuable (2015, 310). What is intrinsically valuable is the realization of particular kinds of relations, especially narrative relations,\(^{54}\) among temporally separated life events. This is the

\(^{53}\) (2015, 305)

\(^{54}\) Although Dorsey and Velleman both talk extensively about narrative relations, neither says exactly what narrative relations are supposed to be. For present purposes, a rough characterization will suffice. The idea is that a narrative of a life is an accurate, story-like description of the events in the life. And the narrative relations among events are the relations that make the events suitable for description as part of such a narrative.
Relational View about the value of the shape of a life.\textsuperscript{55} For example, consider a life with a steady sequence of low levels of well-being followed by a pronounced incline. If this life’s shape seems valuable to us, this may be because of the kind of life story this shape signifies. This shape might signify a life in which repeated episodes of onerous work lead to the realization of a long-sought goal. And, according to the Relational View, one’s life events being related like this may be actually intrinsically valuable.

Now, note that, according to the Relational View, because it says that a life’s shape is not itself of any intrinsic value, it follows that anyone who takes SLH literally to be a true claim about intrinsic value is mistaken. But the mistake is the minor one of attributing intrinsic value to something that merely accompanies intrinsic value. And, as Dorsey argues compellingly, the Relational View has some clear advantages over its competitors (2015, 315-322). So, we do not yet have cause for concern.

Having argued for the Relational View, Dorsey sets out to show us that the significance of the shape of a life does not threaten aggregationism after all. But it is not the Relational View itself that saves aggregationism.\textsuperscript{56} Saving aggregationism requires that the person’s levels of synchronic well-being reflect the value of diachronic features of the life. That is, it requires accepting that the valuable relations among temporally separated life events can affect a person’s synchronic well-being at the moments when these events occur.\textsuperscript{57} This, of course, makes synchronic well-being diachronically sensitive.

\textsuperscript{56} In fact, Dorsey interprets Velleman as accepting the Relational View and rejecting aggregationism because of it (Dorsey 2015, 310-313).
\textsuperscript{57} See Dorsey (2015, 325).
The problem is that if synchronic well-being is diachronically sensitive (in particular, future-sensitive) to the extent required to preserve aggregationism, then it turns out that SLH’s “sequence of good and bad times,” which the Relational View is intended to explain, is not what we imagined it to be. Returning to the example just broached, the Relational View says that the period of consistently low well-being ending with a sharp incline may indicate a valuable relation among the events of that period. But, due to future-sensitivity, the life may not actually have that particular shape; there may not be a period of consistently low well-being after all. Hence, on Dorsey’s proposal, we are mistaken, not just about whether the shape of a life is intrinsically valuable, but about the shape itself. This is the distortion problem. Thus, future-sensitivity seems to undercut our reason for considering the significance of a life’s shape in the first place.

A possible response here would be to suggest that, when thinking about the shape of a person’s life, we attend to what seem to be the person’s levels of synchronic well-being. And if it turns out that, due to future-sensitivity, the actual levels of synchronic well-being are different from what they seemed to be, then that is acceptable. According to this response, we should not necessarily expect our actual levels of well-being to be exactly as they seemed. After all, if a philosophical theory tells us that reality is a bit different from the way it appears, this does not necessarily amount to a “distortion” in any problematic sense.

According to this response, when thinking about how a person’s life is going, we tend to be attuned to something other than the sequence of genuine synchronic well-being levels. Perhaps we tend to focus on levels of pleasure or happiness (which might be just
symptoms, or maybe components, of synchronic well-being proper). Or perhaps what we pay attention to is simply the synchronic well-being levels with all of the future-sensitivity factored out (or, put another way, never factored in). Whatever it is we are attuned to, call this the person’s level of apparent well-being. Although Dorsey understands the shape of a life in terms of actual synchronic well-being, suppose we were willing to characterize the shapes of lives in terms of apparent well-being instead. Adjusting the Relational View accordingly, we could still hold that the shape of a life—now understood as the sequence of levels of apparent well-being—may be a sign of valuable diachronic relations among life events. And we could go on to say that the value of these relations is reflected in the person’s levels of actual well-being at the moments when the relevant events took place. Now, as long as apparent well-being is not itself future-sensitive, the distortion problem arises only if we insist that apparent well-being must match actual well-being. But, according to the response here, the two are different, and it is no problem if they come apart.

This may seem fine. However, it amounts to admitting that apparent well-being (whatever it turns out to be) plays a role much like a significant evaluative category in its own right. After all, the suggestion was that apparent well-being is what we typically have been paying attention to in our judgments about how a person is faring during the moments of her life. Of course, even if we did admit an important role for apparent well-being, we could still hold that genuine synchronic well-being is future-sensitive. But then it begins to

---

58 Dorsey clarifies, “[W]hen using the phrase ‘good times’ [in SLH], I simply refer to times during which one is net benefited by particular welfare goods; times at which one has a high level of synchronous welfare” (2015, 305).
look like the primary job of genuine synchronic well-being is just to register diachronic sources of value in a way that preserves aggregationism. 59 This would be a pyrrhic victory for future-sensitivity and aggregationism alike.

If defenders of future-sensitivity are not willing to countenance an important role for apparent well-being and the corresponding reduction in the role for genuine synchronic well-being, then they are stuck with the distortion problem. But, for some, that might be an acceptable cost. One option here would be to suggest that the shape of a life is not all that significant anyway, or to suggest that it is not important that lives have the shapes we took them to have. Then, nonetheless, we might still care about the value of diachronic features of a person’s life. And we might want a conception of synchronic well-being that registers this value. However, future-sensitive synchronic well-being has more serious problems, to which I now turn.

III. The instability problem

As already noted, among the valuable relations that we might think hold among temporally separated events are certain kinds of narrative relations. 60 Although Dorsey and Velleman agree that narrative relations among temporally separated events can affect well-being, 61 they differ with regard to how and when these narrative relations are valuable (or disvaluable) for the person in whose life they occur. Whereas Velleman holds that this value

59 Indeed, Dorsey concludes his discussion of aggregationism by saying, “If we find that [aggregationism] should be accepted, then we will accept [that synchronic well-being is diachronically sensitive]” (2015, 329). But this is not as innocent as he makes it out to be. What he leaves out is that this would yield a notion of synchronic well-being that was not suitable for characterizing the shapes of lives.

60 As I suggested in note 54, we might think of narrative relations among events as those relations that make events suitable to be described as part of a person’s life story.

attaches to stretches of a person’s life (2000, 71), Dorsey proposes that the value could attach to the life at the specific moments when the narrative relations are partially realized (2015, 325-327). In other words, Velleman holds that narrative features affect only diachronic well-being, while Dorsey suggests that these features can affect synchronic well-being (and then affect diachronic well-being only derivatively).

The problem I will raise now ultimately depends on the observation that the story of a person’s life involves how well the person was faring at different points in her life. In other words, a person’s levels of synchronic well-being affect the narrative. The problem is that if synchronic well-being registers all the value of a life’s narrative features, then the determinants of synchronic well-being may include those very levels of synchronic well-being. The result, as I will show, is an inordinate degree of complexity.

To see the contours of the problem, it will be helpful to look at how it would arise for a theory that does not assign narrative a special role (and so differs from Dorsey’s in that regard). Consider a view of well-being that holds that SLH actually expresses a general truth about what is good for us. We are setting aside our original concern with explaining our SLH judgments; now we are supposing that SLH is a substantive evaluative truth. The view says that the sequence of synchronic well-being levels itself can be intrinsically valuable.62 Suppose that the view we are considering also says that synchronic well-being is diachronically sensitive in such a way that important diachronic features in a life (such as its particular sequence of synchronic well-being levels), insofar as they affect the goodness of the life at all, affect it by way of the person’s level of synchronic well-being at the moments

62 As such it would be an instance of the view Dorsey labels the Intrinsic View, in contrast to his preferred Relational View (2015, 314). The Intrinsic View is advanced by Glasgow (2013).
in which these diachronic features are partially realized. Crucially, the diachronic sensitivity here includes future-sensitivity. In sum, we are imagining a view that conjoins acceptance of SLH as a genuine truth with the main structural feature of Dorsey’s suggested view of synchronic well-being. According to the imagined view, the sequence of levels of synchronic well-being across an entire life could affect each of those very levels, thus affecting the sequence, thus affecting the levels, and so on. The well-being levels exhibit a sort of instability; so, call this the instability problem. In short, the general problem is this: *Future-sensitivity allows synchronic well-being to be sensitive to factors that themselves depend on synchronic well-being.*

Just how much of a problem is the instability problem? Well, if the recursive determination of the particular well-being levels and the value of the sequence of those levels does not bottom out or otherwise stabilize, then a person’s level of synchronic well-being at a time will be indeterminate. That, it seems to me, would be highly problematic. But even if we do not end up with indeterminacy, the instability problem shows that, given future-sensitivity, how well a person is faring at a time is a considerably more complex matter than what I suppose any of us had been imagining. Indeed, I think it is a level of complexity extreme enough to make the position that entails it unacceptable. To see the extent of the complexity, we can continue working with the view just sketched and look more closely at what, according to this sort of view, determines a person’s well-being at a moment.

We are focusing on a person’s level of well-being at a particular moment. Due to future-sensitivity, the person’s well-being at that moment may depend on events that occur
any time later in her life. To avoid additional complications, let us assume that all those subsequent events are determinate. Intuitively, we are taking a postmortem perspective relative to the life in question.

Now, on the view we are working with, SLH is a substantive truth; i.e., the shape of a life is itself valuable. So, let us assume that there is some function (perhaps just a partial function) that maps life shapes, i.e., sequences of synchronic well-being levels, to degrees of goodness. At this juncture, just for now, let us also assume that the levels of well-being at all the other points (both past and future), besides the moment in question, are fixed. Then we can think of the goodness of the life’s shape as a function of just the level of well-being at the particular moment in question.

Turning our attention to what determines the level of well-being at that moment, suppose that all synchronic determinants of value are already accounted for. We are also supposing that diachronic features of the life, like the life’s shape, affect the well-being at that moment. For simplicity, suppose that the life’s shape is the only diachronic source of value in play. Then we can understand the level of well-being at the moment as a function of just the goodness of the life’s shape.

Now we can put these two functions together. Start with the person’s level of apparent (i.e., without the effects of future-sensitivity) synchronic well-being at the moment. Plugging this value into the first function yields a value for the goodness of the life’s shape. Then plug the goodness of the life’s shape into our function for the well-being at the moment. This gives us an adjusted level of synchronic well-being. From there we calculate a new value for the goodness of the life’s shape. And so on. If this process of
recursive evaluation stabilizes, then we will get a determinate level of well-being at the moment, along with a determinate goodness of the curve. So far, this does not seem so bad. The picture may appear more attractive when we notice that our two functions resemble a simple system of equations and that the recursive process of evaluation may not be the only way to reach a solution. If the system of equations can be solved directly and has a unique solution, then the situation here does not seem so complicated after all.

However, so far our illustration has relied on a crucial assumption that keeps complexity at bay, an assumption that would never hold. The assumption was that, other than at the point in question, the levels of well-being at all the other times are fixed. But, according to the view we are looking at, these levels would not be not fixed; they, too, are diachronically sensitive. Hence, when we get a new value for the goodness of the life’s shape, this affects (at least potentially) the level of synchronic well-being at every other moment in the life. And, of course, the adjustment of these well-being levels yields a new shape. So, not only do we have interdependence between well-being at the moment and the life’s shape; we have interdependence among the levels of well-being at all the moments of the life.

To rein in this additional complexity, we might hope to avail ourselves of the shortcut mentioned before, of bypassing all the recursive reevaluation by settling all the evaluations in one fell swoop, like solving a system of equations. But note that it would no longer be a system of equations in just two variables; it would be a system of equations in as many variables as the number of discrete moments of synchronic well-being in the person’s life. Would such a system of equations have a solution? Maybe. Would it be unique?
Maybe. But even setting aside worries about indeterminacy, this picture shows the determinants of synchronic well-being to be far beyond what we could possibly grapple with in anything like our ordinary thinking about how a person is faring at a particular point in time. Hence, whether we focus on the instability we find in the recursive process of adjusting all the well-being levels, or we think in terms of the massive system of interdependent evaluations, we are confronted with inordinate complexity.

The view Dorsey proposes is subject to a version of the instability problem, but the way the problem arises is a little more complicated. After all, Dorsey does not think that SLH is literally true—at least not if the value in question is intrinsic value. Recall that, according to the Relational View, the shape of a life is merely a signifier of the presence of something else which is genuinely intrinsically valuable, viz., narrative relations among the events in the life. So, we do not have here the tight circuit in which the value of the sequence of synchronic well-being levels affects those very levels, and so we have not yet run into the instability problem.

But it turns out we have only deferred the problem, not averted it. Yes, we can say that the events and their narrative relations are what affect synchronic well-being levels. The trouble is that sometimes a person’s levels of synchronic well-being are important elements in the narrative. Put another way, it matters to the story of a person’s life whether and when she was faring well during that life. Specifically, a rise or fall in how well the person was faring may constitute an important narrative link between events. For example, consider a person who is laid off from a job she had for many years. Imagine that she soon finds a new job, and a year into that new job she has a much higher level of well-being than
she had while in her previous job. The details here could be filled in in various ways, but, regardless, the person’s increased level of well-being will matter to her story. Changes in a person’s well-being level over time may constitute narrative links between different parts of that person’s life. According to Dorsey’s proposal, the value of any diachronic narrative relations like these will be reflected in a person’s synchronic well-being at the various moments at which the relations are partially realized. So, we end up with this: narrative relations affect levels of synchronic well-being, and levels of well-being partially constitute narrative relations. The reciprocal dependence of narrative relations and synchronic well-being levels realizes the recursive structure that is the signature of the instability problem.

For a model of this version of the instability problem, we can look to one of Dorsey’s focal examples: the life of O. J. Simpson. Simpson had great success as a collegiate and professional football player, followed by a series of fairly public disasters. These included the murder of his ex-wife, for which many people blamed Simpson himself, and convictions and imprisonment for robbery and kidnapping. Relatively early in his life, Simpson seemed to be very faring very well. One valuable event early in his life was winning the Heisman Trophy (college football’s highest individual honor). According to Dorsey’s view, Simpson’s apparently negative trajectory after that point is a symptom, a sign, of the undesirable narrative that comprises many of the events in Simpson’s life. And Dorsey holds that the events and their unfortunate narrative may be reflected in Simpson’s levels of synchronic well-being at the various moments of his life. This is how Dorsey explains it:

63 Though I do not know enough about Simpson’s life to estimate Simpson’s level of well-being, I will stick with Dorsey’s assessments for the sake of exposition.
The shape of a life continues to maintain the relevant signatory significance even if the relations borne by the relevant events can help to determine their synchronic value. For instance, consider a typically valuable event in a life, such as winning the Heisman Trophy. This event, surely important for synchronic well-being, could be a stepping stone, or merely a teaser: early success followed by a disastrous and shameful downfall. If it is a stepping stone to future success, one might say that the synchronic value of this event is greater than the synchronic value of an identical event which is just a teaser. For O.J. Simpson, his Heisman victory was, in fact, less synchronically valuable given what we now know: that it helped to frame his his ever-so-public downfall. (2015, 325-326)

Whether this is a case that manifests the instability problem depends on whether the valuable (or disvaluable) narrative features of Simpson’s life themselves depend at all on his levels of synchronic well-being. And we should think that they do. The most significant narrative feature here is Simpson’s downfall. Although many of the events that punctuate the downfall can be characterized without explicit reference to Simpson’s well-being—for instance, the murder of his ex-wife, legal defeats, loss of wealth, etc.—Simpson’s declining well-being is an essential element in the story. To see this, imagine that Simpson’s well-being did not decline, but that it has remained fairly constant over his life. Without declining well-being over the course of his life, the other elements of the story would not constitute the same sort of downfall, and the story would not be so tragic (or, at least, it would not be tragic in the same way).

We can see the instability problem when we ask, exactly how well was Simpson faring at the time he won the Heisman? If we ignored subsequent events, we would likely say that Simpson was faring quite well at that point. But, due to future-sensitivity, we must evaluate that early period from the standpoint of subsequent events, including, of course, Simpson’s downfall. Taking the downfall into account yields a reduction in Simpson’s Heisman-era
well-being level, compared to an evaluation of that period that does not register Simpson’s eventual misfortune. But this requires a reconsideration of Simpson’s narrative, because it has turned out that Simpson did not have such a high early level of well-being to frame his decline. Since Simpson’s decline is not as great, the narrative is not as tragic, and this adjusts back upward, to some extent, his level of well-being at the time he won the Heisman. Of course, this also re-elevates the severity of Simpson’s decline, which worsens the narrative and reduces his early well-being, thus reducing again the severity of decline. And so on.

We have caught just a glimpse of the complexity here. So far we have just focused on Simpson’s level of well-being at a moment (or, perhaps, a brief period) in his life. A full accounting could not ignore his well-being at all the other other times. That is because, on

---

64 Does this really make the life less tragic? Certainly, it depends on what constitutes tragedy. If, as I have been assuming, a great decline in well-being may contribute to a tragic life story, then it is plausible that, as I am suggesting, a mitigated decline may make the narrative less tragic. On the other hand, in the present case, the mitigation of the decline is due to the initial level of well-being turning out to be not as high as it appeared at the time. This disparity between Simpson’s level of apparent well-being and his level of genuine well-being seems tragic in a very different way: it is tragic that he was actually faring worse than he (and everyone else) thought. So, do the effects of future-sensitivity on Simpson’s Heisman-era well-being level improve or worsen the narrative? It is difficult to say. However, either answer would suit my present purposes. According to either answer, Simpson’s level of genuine synchronic well-being is playing a role in the narrative. And if that level changes, it also changes the narrative and its value. Then this change in the value of the narrative features of Simpson’s life would be reflected in his well-being at the moment in question. Thus, we end up with interdependent evaluations either way.

65 Dorsey might respond here by urging that it is not only Simpson’s Heisman-era well-being that is shifted lower by subsequent events; his well-being levels at subsequent moments in the narrative are shifted downward as well (2015, 236, note 36). And this would preserve the precipitousness of Simpson’s decline. I think this is a reasonable suggestion, but this would not avert the instability problem. It is not just relative levels of synchronic well-being that matter for narratives; absolute levels matter, too. As I remarked in note 52, the story of a life that goes from very bad to pretty good is quite different from a story that goes from pretty good to really great. The relevant point regarding Simpson’s life is that a story that goes from good to very bad is tragic in a different way from a life story that goes from middling to very, very bad. So, a uniform downward shift of well-being levels would still change the narrative and, perhaps, its value. And any such change in value would be reflected by changes in the well-being levels at points during the narrative. This, of course, affects the narrative again, and so on.

84
the view Dorsey proposes, a change in the value of the narrative may affect the level of synchronic well-being at any time encompassed by that narrative. So, insofar as these other levels of well-being themselves figure in significant narrative relations, we will have many parallel instances of this recursive evaluation process. Or, as before, we could think of the interrelated evaluations across Simpson’s life as a massive system of equations. Either way, the complexity is extreme. Whether we end up with determinate levels of synchronic well-being will depend on the details of the theory—specifically how the value of the narrative is supposed to be reflected in the momentary well-being levels. But even without indeterminacy, the sweeping interdependence of the scores of evaluations involved here is not plausible. Hence, we should reject Dorsey’s proposal that synchronic well-being may be future sensitive in a way that would allow it to reflect the significance of diachronic narrative relations.

At this point, it is worth noting that the instability problem is not a totally unavoidable consequence of future-sensitivity. In principle, the instability problem could be blocked by limiting the sorts of future facts to which present synchronic well-being is sensitive. In general, the instability problem affects views that allow that a person’s present synchronic well-being may depend on some future facts that themselves depend on her present synchronic well-being. As we saw at the beginning of this section, the instability problem affects the view that says the shape of a life is intrinsically valuable and that its

---

66 I would expect that, at this point, at least for understanding a life’s narrative features, focusing on apparent well-being instead of genuine (future-sensitive) synchronic well-being may look attractive again. The idea would be that, in life stories, some non-future-sensitive evaluative category tends to be most relevant. This is to return to the response to the distortion problem considered at the end of the preceding section. Now, as then, it would amount to an admission that genuine synchronic well-being is not as important to us as we had thought.
value is to be reflected in synchronic well-being levels. We have just seen that the problem affects the view Dorsey proposes, because he holds that synchronic well-being may be affected by narrative features of a life, and these features may involve levels of synchronic well-being. What sort of view would not be affected? One might think that a simple desire-satisfaction view would be immune. Consider a desire-satisfaction view of synchronic well-being that says that a person’s present synchronic well-being is increased if one of her present desires is satisfied at any point in the future. The instability problem will arise even for this sort view. Suppose a person presently desires to be better off in the future than she is now. Or consider a parent who wants her children to fare better than she is faring. Then the person’s present synchronic well-being will depend on facts that themselves depend on her present synchronic well-being, and we have the interdependence that generates the instability problem. So, though it is possible, in principle, to block the instability problem, most views that countenance any significant future-sensitivity will not.

IV. Problems for past-sensitivity?

I remarked in Section I that I do not want to rule out the possibility that synchronic well-being is past-sensitive. We can ask: Are the problems just described problems for diachronic sensitivity in general, or are they due to future-sensitivity in particular? I will argue that the problems are far worse with future-sensitivity than with just past-sensitivity.

67 Dorsey (2013) advances a version of desire satisfactionism like this. See also Sarch (2013).
68 One might think that, even though we find the same recursive structure, perhaps the instability problem will not generate all that much complexity for this sort of desire-satisfaction view. To temper any such optimism, imagine a family tree full of people each of whom wants her children to be better off than she is. The evaluations of synchronic well-being become very messy very quickly.
69 This may be an understatement. I, for one, am not aware of anyone who advances future-sensitive synchronic well-being and avoids the instability problem.
First consider the distortion problem. We encountered the problem when future-sensitive synchronic well-being was supposed to reflect the value of the diachronic features of a life in a way that might preserve aggregationism. But it is hard to imagine someone trying seriously to use past-sensitivity to preserve aggregationism. To see why past-sensitivity is not up to the task, imagine a life that, intuitively, was going well until near the very end. And suppose at the end there is some event that stands in very undesirable relations to the life’s earlier events. Intuitively, it tarnishes much of what came before. But if we have past-sensitivity without future-sensitivity, all of the disvalue of this event must be accounted for at the time of the event and thereafter. Now, since the event took place near the very end of the life, there is very little time for the person’s synchronic well-being to account for all the negative value. If we are trying to preserve aggregationism, then the steepening of the downward well-being curve must be extreme. Essentially, to make the sum come out right, the final, short period of the life would have to be a period during which the person’s well-being level was far, far below zero. But how badly someone is faring at a time does not depend on the amount of preceding goodness that needs to be compensated for at that time. Such an evaluation of a person’s level of well-being would have nothing to support it except that it is what we need in order to make the numbers come out the way aggregationism predicts. Assuming that well-being levels are not ruled by bookkeeping considerations like this, past-sensitivity is an inappropriate tool for preserving aggregationism. Since past-sensitivity is not to be invoked for this purpose, invoking it for this purpose cannot be the basis of an objection to it.
There is also a simpler, more fundamental reason that past-sensitive synchronic well-being does not generate the distortion problem. The distortion in the distortion problem was the difference between what we took to be the shape of a life what the shape turned out to be once diachronic factors were reflected in synchronic well-being. This difference is likely to be much less drastic when the diachronic features to be taken into account are limited to those that occurred in the past. After all, it is more likely that we already had been taking many of those factors into account (to whatever extent they appeared relevant), simply because they would have happened already. This shows us a way in which a conception of synchronic well-being that countenances past-sensitivity is unlikely to be as revisionary as one that allows future-sensitivity.

Now consider the instability problem. The issues here are a little more complex. I will end up claiming that, though past-sensitivity does present complications somewhat similar to those we find with future-sensitivity, they are not as severe. Regarding whether the complications that apply specifically to past-sensitivity are worrisome enough to discredit it, I will withhold judgment.

Suppose synchronic well-being reflects valuable diachronic narrative relations in a person’s life via past-sensitivity. Imagine a case in which a significant period of a person’s life culminates at some moment with an achievement that brings a satisfying narrative unity (however understood) to that period. And suppose that a significant element of the narrative is the person’s heightened level of synchronic well-being at the end of that period. How do we evaluate the person’s level of synchronic well-being at the endpoint? If we suppose that the higher the level of well-being at the endpoint, the better the narrative, and
if we are supposing that this diachronic benefit is reflected in synchronic well-being at the endpoint, then we have something like a positive feedback loop: The higher the level of well-being, the better the narrative, the higher the level of well-being, and so on. (It is not hard to imagine cases where the level is recursively pushed downward or cases that manifest an up-and-down oscillation.) I do find this worrisome.

But, compared to the situation with future sensitivity, there is an important difference. With past-sensitivity, we have action only at one point on the timeline. We are assessing the level of well-being at the latest point in the timeline, and the evaluation has a recursive character because the well-being level depends on the narrative, and the narrative involves facts at that moment at the end of the timeline, potentially including the level of well-being. So, though the narrative varies as the level of well-being varies, the level of well-being varies at only one moment.

Compare this sort of point-instability, if you like, to the diachronic instability that comes with future-sensitivity. Consider some event that occurs at the end of some period in a life. The outcome there may, due to future-sensitivity and narrative relations between present and past, affect the person's well-being level at some earlier time. This retroactive effect on the well-being of the person may make a difference to the narrative that encompasses the time of this effect. As the narrative changes, we get changes to the levels of well-being at other times encompassed by the narrative. As these well-being levels

---

70 Here I am assuming that along with future-sensitivity, according to the sort of view we are imagining, we have also at least some degree of past-sensitivity. That is because it is implausible to think that some narrative features affect how well someone is faring but that past parts of the narrative do not do so at all. It is of the nature of narrative that the significance of later parts of the narrative depends, to some extent, on what came before.
change and insofar as the narrative involves them, there are additional changes to the narrative. And so on. This differs from the purely past-sensitive case in which the only moving piece was the well-being level at the endpoint of the narrative.

We might illustrate how past-sensitivity differs from future-sensitivity with respect to the instability problem as follows: With the value of narrative relations captured through future-sensitivity, the unfolding of a present event initiates cycles of retroactive alteration of the preceding pieces of the narrative. With the value of narrative relations captured through just past-sensitivity, the unfolding of a present event initiates, at most, a cycle of reevaluation of just the present. If the reevaluation extended to preceding well-being levels, that would be future-sensitivity, not just past-sensitivity.

Hence, the case I have made against future-sensitive synchronic well-being does not appear to pose such severe problems for past-sensitivity.
Chapter 4: Desirability as a Kind of Egocentric Value

Introduction

Going on walks is good for me. So is gardening. So is reading good philosophy. I will say that each of these things is desirable for me. In calling something desirable for me, I do not mean that it increases my well-being or improves my life, even ceteris paribus. The goal of this chapter is to advance the thesis that desirability is a distinctive sort of egocentric value that is not analyzable in terms of its contribution to well-being or the goodness of a life.

A methodological obstacle stands in the way of this goal. What would seem to be the most straightforward approach is unavailable. This approach would be to start from a large collection of examples of things that are clearly desirable for me, and ask what they all have in common. From there we might hope to build a definition of desirability. Then, with a definition in hand, we could see whether and to what extent desirability is different from well-being and life goodness.

The problem with this straightforward approach is that, since desirability is seldom recognized as a distinct sort of value, we do not have ready to hand a definitive list of paradigm examples. I expect that many readers will have read the initial examples of what is desirable for me—going on walks, gardening, and reading good philosophy—as examples of things that increase my well-being or make my life better. Hence, these examples are not
particularly helpful for distinguishing desirability from these other sorts of egocentric value. Furthermore, if I were to identify examples of desirable things that clearly do not, even *ceteris paribus*, increase my well-being or improve my life, this would be unlikely to help my cause either. That is because advocates of the position I am opposing tend to assume that all egocentric value should be understood in terms of well-being or life goodness. So, to the extent that an example is not something that contributes to my well-being or the goodness of my life, the predictable response from the opposing camp will be to conclude that the example is not an example of egocentric value at all.

An alternative approach would be for me to introduce desirability by stipulating a definition for it. Then I could use this definition to demonstrate that desirability was distinct from well-being and life goodness. But this is not very attractive. It would amount to starting with a theory, without first identifying what the theory is supposed to be a theory of. This *theory-first* approach, as we might describe it, sounds perverse. One immediate problem is that it is trivial to define some property that is distinct from and not-reducible to some other given properties. Even if the definition gives us a notion of a property that looks broadly evaluative and has some other neat features, it would not be of any obvious philosophical interest.

However, the theory-first approach is, in effect, the approach I will adopt here. That is to say, my development of the notion of desirability, and the ensuing arguments about it, will not be grounded in an independently circumscribed set of examples. To mitigate the perversity of this and bolster the philosophical significance of the results, I will offer some theoretical context before defining desirability. Specifically, I will present and
discuss a well-known conception of non-egocentric value, and use this as a sort of anchor for our conception of desirability as a kind of egocentric value.

I will begin with a famous and, I will assume, quite plausible conception of non-egocentric value adumbrated by Isaiah Berlin. My guiding thought will be that desirability is the egocentric analog of the sort of non-egocentric value described by Berlin. However, Berlin can take us only so far. Berlin’s remarks about value are suggestive and indicate the broad contours of an important conception of value, but it is not detailed enough to guide us toward a theory of desirability. To further develop the sort of conception of value Berlin offers, I will draw extensively on David Lewis’s dispositional theory of value. This is crucial: The point of discussing Lewis’s theory is not to defend a dispositional theory of value or Lewis’s way of developing it. The goal is much more modest; I discuss Lewis’s theory as a way of enriching the conception of value we find in Berlin. I will argue that a dispositional theory like the one Lewis offers is a theory of the sort of value described by Berlin. But this will not be an argument that the dispositional theory is completely adequate, or that the dispositional theory is the only defensible metaethical theory of this sort of value.

71 Hence, for someone who does not find Berlin’s picture intuitive or attractive, my way of developing desirability is unlikely to be very compelling. I readily admit this. The purpose of this chapter is not to prove that desirability is an indispensable part of our evaluative thinking. Rather, the goal is to work from a conception of non-egocentric value that many will find familiar and agreeable, and show how reflection on this sort of value opens up theoretical space for desirability as a sort of egocentric value. Inevitably, being persuaded by such a project depends to a large extent on being persuaded of its starting points.

72 I am arguing Lewis’s theory is a theory of the kind of value Berlin has in mind. Might an alternative theory of this sort of value serve my purposes equally well? Possibly. Suppose a fitting attitude account of value could capture the distinctive features of Berlin’s view. If the fitting attitude analysis can also be adapted in a way analogous to the adaptation of the Lewisian theory described in Part II—restricting the scope to essentially egocentric contents, instead of non-egocentric contents—then this might be another acceptable path to a conception of desirability. One reason for using the Lewisian dispositional view in particular is its similarities to Sidgwick’s view of desirability.
Two distinctive features of Berlin’s conception of value are the way in which it is pluralistic and the sense in which it is intersubjective. I will show how a broadly Lewisian dispositional theory of value can help explain why value would have these distinctive features. Once we see how the dispositional theory provides an attractive articulation of Berlin’s conception of value, then we can investigate how the theory might be adapted or extended. The key observation is that the dispositional theory can yield either a non-egocentric or an egocentric theory of value, depending on how we specify one of its parameters. With one specification, we get a notion of value like Berlin’s. Specified another way, we get a distinctive notion of egocentric value, which I call desirability. That is how I aim to define desirability.

Desirability, thus conceived, inherits analogs of the distinctive features of Berlin’s notion of non-egocentric value. And we can notice that something like desirability already plays a familiar role in our evaluative thinking. Furthermore, it is straightforward to show that desirability is distinct from well-being and life goodness, in the sense that it does not stand in the contribution relation to either. In sum, what we find is that the category of desirability collects certain things that are valuable for me—like taking walks, gardening, and reading good philosophy—in a unified class which, although related to my well-being and life goodness, is not easily understood in terms of either.

Part I of this chapter introduces Berlin’s conception of value and presents a dispositional theory, like Lewis’s, that can explain distinctive features of this sort of value. Part II is intended to be more constructive than analytical. It adapts the theory in Part I to give us a view of desirability, and I draw out interesting implications of this view.
Part I: Non-egocentric value

I.1 Berlin’s conception of value

In his essay, “The Pursuit of the Ideal,” Berlin remarks, “Values may easily clash within the breast of a single individual; and it does not follow that, if they do, some must be true and others false” (2013, 12). Shortly thereafter, he elaborates

These collisions of values are of the essence of what they are and what we are. If we are told that these contradictions will be solved in some perfect world in which all good things can be harmonised in principle, then we must answer, to those who say this, that the meanings they attach to the names which for us denote the conflicting values are not ours. We must say that the world in which what we see as incompatible values are not in conflict is a world altogether beyond our ken; that principles which are harmonised in this other world are not the principles with which, in our daily lives, we are acquainted; if they are transformed, it is into conceptions not known to us on earth. But it is on earth that we live, and it is here that we must believe and act. 73

Berlin offers us a picture according to which values come into conflict, for a single person and between persons, but are still values in spite of this. This is Berlin’s value pluralism, which he never presented systematically, but for which he is nevertheless famous. In a later essay, in which he is describing the evolution of his moral and political thinking, Berlin describes the limits to his pluralism:

I am not a relativist; I do not say ‘I like my coffee with milk and you like it without; I am in favour of kindness and you prefer concentration camps’ – each of us with his own values, which cannot be overcome or integrated. This I believe to be false. But I do believe that there is a plurality of values which men can and do seek, and that these values differ. There is not an infinity of them: the number of human values, of values which I can pursue while maintaining my

73 (2013, 13-14)
human semblance, my human character, is finite – let us say 74, or perhaps 122, or 26, but finite, whatever it may be. And the difference this makes is that if a man pursues one of these values, I, who do not, am able to understand why he pursues it or what it would be like, in his circumstances, for me to be induced to pursue it. Hence the possibility of human understanding.\(^7^4\)

Berlin’s value pluralism is not an anything-goes relativism. Though there are many values we might pursue, values which might even stand in opposition to each other, it is not the case that just anything that anyone values counts as a value.\(^7^5\) Our common humanity limits the range of plurality. Furthermore, due to our common humanity, when we value these human values, our ways of valuing are intelligible to others, even if they value differently.

In the rest of Part I of this chapter, I will focus on the features of Berlin’s conception of value just highlighted. To summarize: First, it is a conception of value according to which there are many values, which may not be realizable or even pursued together. This is the pluralistic aspect. Second, this conception of value allows for a kind of intersubjectivity. Values, despite their diversity, are such that beings like us can value them, and, when we do value them, we make ourselves intelligible to one another.

I.2 A dispositional theory of value and imaginativist idealization

I will now attempt to show how a dispositional theory of value, of the sort offered by David Lewis, can illuminate and further articulate the conception of value which I have been

---

\(^7^4\) (2000, 11-12)

\(^7^5\) Following Berlin’s usage, when I talk about values, I am talking about what is of value. This is very different from talking, as people frequently do, about a person’s values. A person’s values are what she actually values, those things to which she has a valuing attitude, not all of which may be values in Berlin’s sense. For a similar statement of this distinction see Svavarsdóttir (2014, 85).
describing. A dispositional theory of value understands values as that which we are disposed to value. But we need to be more specific than this; sometimes we value things that are not valuable and fail to value things that are. The key, then, is to say under what conditions valuing entails value. We can follow Lewis (1989, 113) in taking a dispositional theory of value to have this form:

Something of the appropriate category is a value iff we would be disposed, under ideal conditions, to value it.

We articulate the theory by saying something more specific about the italicized parts of the definition. To wit: the category of the items which could be values, the extension of ‘we’, the specification of the ideal conditions, and what it is to value something. For the first part of this chapter, we can take category of valuable things to be ordinary propositions. (An adjustment to that part of the theory will be the key difference in Part II of this chapter.) I will have a bit more to say about who is included in ‘we’ in Section I.4 below. The ideal conditions will be our primary focus for the rest of Part I of this chapter.

Before considering the ideal conditions, I need to make a few comments about the valuing attitude. After ruling out kinds of feelings and beliefs, Lewis settles on the thought that valuing must be a kind of desiring. However, examples like that of an unwilling addict, who desires to take her drug but does not value taking it, suggest that what a person values may diverge from what she desires. And so Lewis concludes that valuing something is best thought of as having a second-order desire for it, a desire to desire it (1989, 114). This

76 Lewis also notes another question is the modal status of the statement, but I will not have anything to say on this issue.
argument is a little too quick to be convincing. It seems to me that it is open to us to say that an addict, even an unwilling one, insofar as she genuinely desires to take the drug, values taking the drug. Of course, she may also value abstaining. Perhaps she values taking the drug for the high, but values abstaining (or even disvalues taking the drug) because of the cognitive impairment. Furthermore, I worry that higher-order attitudes are inappropriate for the dispositional theory of value simply because I suspect that, as a matter of psychological fact, our dispositions to have higher-order attitudes are too spotty. I do not think I have particularly robust dispositions to have higher-order attitudes. Even less do I think it likely that many such dispositions will be widely shared, let alone shared across the great majority of humanity (which, as we will see below, is what we would need for the dispositional theory to hit our target). Ultimately, I wish to remain noncommittal about exactly which among our attitudes we should think of as attitudes of valuing. So, when I talk about the attitude of valuing I will have in mind the class of attitudes that meet at least a minimal sufficient condition for counting as valuing in any way. For purposes of this chapter, in this class I hesitantly include all the common pro-attitudes of desiring, liking, appreciating, caring about, being in favor of, etc.  

Svavarsdóttir (2014, 94-95) argues that valuing is a distinctive sort of pro-attitude. One of its distinguishing features is that it grounds emotional, cognitive, and deliberative dispositions that are stable across time. Might such a conception of valuing be appropriate for purposes here? The idea is an attractive one, and, to my mind, much more attractive than Lewis’s own second-order desire proposal, but I wish to remain neutral. To account for the features of value—pluralism and intersubjectivity—that are important for my purposes, the dispositional theory’s ideal conditions are much more important than the particular features of the valuing attitude. Of course, if it could be shown that invoking Svavarsdóttir’s view of valuing within a Lewisian dispositional theory yielded a notion of value more like Berlin’s, then I would happily accept it.
With those preliminaries noted, we can focus on the ideal conditions. According to Lewis, the ideal conditions are those of “full imaginative acquaintance” with the potentially valuable object (1989, 121-126). For a person to have imaginative acquaintance with things that might be valuable is for the person to “imagine vividly and thoroughly how it would be if these putative values were realised” (1989, 121). Full imaginative acquaintance is simply the maximum degree of imaginative acquaintance that can be achieved without altering the relatively permanent features of her psychology, like her dispositions to desire. I will say that these ideal conditions are imaginativist.\(^78\)

Imaginativist ideal conditions differ in several crucial ways from conditions which give the responder full information. As Lewis stresses (1989, 125-125), full imaginative acquaintance has nothing to do with extra information. The fullness of imaginative acquaintance is a matter of achieving the maximum level of awareness with the maximum vividness of the proposition being evaluated, not knowing everything there is to know about it. That is not to say that information plays no role in the process of reaching evaluations. In cases in which a putative value is put up for examination, but in an overly general or

---

\(^78\) Though Lewis invariably characterizes this state in terms of imagination, I worry that this is slightly misleading. Reference to the imagination—as a creative faculty with the ability to simulate unexperienced situations—is not essential for characterizing the state under which the disposition to value is supposed to entail value. What clearly matters to Lewis (and I believe rightly so) is that the relevant state of mental acquaintance with the purported value is very vivid; there is a high degree of awareness, presence to mind, of the thing being evaluated. In principle, it does not matter whether this sort of vivid, highly aware acquaintance was achieved by some act of imagination or otherwise (perhaps by memory or present observation). Where imagination is indeed vital is in the epistemology of value, since using one’s imagination tends to be a convenient way (if not the only way) to achieve (or at least approximate) the relevant sort of acquaintance with most putative values. With this caveat noted, I will continue to call the relevant idealized state ‘full imaginative acquaintance’. This has not only the advantage of preserving Lewis’s terminology; it also serves as a reminder of the role imagination typically plays in getting oneself in the relevant state.
otherwise imprecise way, extra information may be required to complete the specification. This process of *refining*—by conditionalizing, conjoining, etc.—the proposition to be evaluated is prior (both as a condition and temporally) to full imaginative acquaintance. Getting extra information may be a step toward specifying the proposition with which we are to be fully aware.

For example, suppose that after a run-in with my libertarian cousin, I find myself preoccupied by a nebulous worry about whether or not free, unregulated commercial trade is valuable. Before I can hope to reach some evaluation, I need more clarity about what I am trying to evaluate. Do I mean free trade at any place and time in the history of human civilization? Or do I care about a particular place and time? If I find my concern is the more specific one—say it is about commerce in economies like that of my region—then, in order to figure out what I am evaluating, I may want to learn a bit about such economies. Suppose I learn that economies like mine have characteristics X, Y, and Z. Now I have something to evaluate. I can ask about the value of there being free, unregulated commercial trade in economies with characteristics X, Y, and Z. The answer will depend on how we are disposed to respond under conditions of full imaginative acquaintance with

---

79 What if what I come to believe about my economy is not true? Suppose I mean to be assessing the value of free market practices in economies like mine, and I read that my economy has characteristics X, Y, and Z. But suppose also that it does not actually have these characteristics. If under conditions of full imaginative acquaintance, we value economies with X, Y, and Z also being free market, does that mean it is valuable for my economy to be free market? The answer has to be *no*. What the valuing disposition here entails is that free market practices *would* be valuable in my economy *if* my economy had X, Y, and Z. This highlights a way in which full acquaintance is very different from full information.
that. Notice that, on this picture, all the additional information already has figured in prior to the imaginativist idealization.\(^\text{80}\)

With imaginativist idealization, the ideal conditions may vary a lot, depending on what is being evaluated. That is because full imaginative acquaintance with one sort of thing, e.g., humane conditions for raising livestock, may be a very different psychological state from full imaginative acquaintance with something else, e.g., respectful political discourse. So, we can say that imaginativist ideal conditions are *object-oriented*; they depend on the object (i.e., the proposition) to be evaluated. This makes the idealization *schematic*, in a certain sense. We cannot give a complete specification of what it is to be in the ideal state independently of the object being evaluated. As Lewis puts it, “On the present theory, when I say that X is a value iff we are disposed to value X under ideal conditions, I do not mean conditions that are ideal *simpliciter*, but rather conditions that are ideal *for X*” (1989, 126).

Another consequence of imaginativist idealization is that the ideal conditions do not take into account the actual situation, the actual practical circumstances, of the evaluator and the world in which she finds herself. Again, here is Lewis: “If you knew too well how costly or how difficult it was to pursue some value, you might reject the grapes as sour, even when imaginative acquaintance with the value itself would have caused you to value it”

\(^\text{80}\) In circumscribing this role for extra information in the evaluative process, I take myself to be elaborating Lewis’s view. However, he is not as clear on this as we might like. Evidence that he accepts the view I am describing is suggested by his statement, “In such cases [in which a putative value is underspecified], a valuer must complete the specification by drawing on his knowledge of the world, else he will not know what he is supposed to imagine” (1989, 125). Still, he is less explicit than I wish to be that any additional information required is involved only *prior to* imaginative acquaintance, in settling precisely what is to be evaluated, as opposed to being part of full imaginative acquaintance.
Note that this is a straightforward consequence of full imaginative acquaintance being a matter of *vivid awareness*, not *full information*. Consider, for example, whether free, on-demand, high quality, universal healthcare is valuable. Surely it is. There is little (if any) disagreement about this. What opponents of universal healthcare proposals and programs complain about are the costs, which are substantial. If a person vividly aware of the operation of a universal healthcare system is also aware of the concomitant costs, then her disposition to value universal healthcare might not be triggered. So, if the proposition to be evaluated was just about universal healthcare and not about the conditions of its implementation and operation, then this would amount to a false negative. To refer to this feature of imaginativism—the exclusion from the state of full acquaintance information about things other than what is being evaluated—I will say that the idealization is *non-situational*.81

An imaginativist idealization, one that is object-oriented and non-situational, yields a dispositional theory of value that is well-suited to capture the pluralism and intersubjectivity characteristic of Berlin’s conception of value. It is to pluralism and intersubjectivity that I will now turn. But, to preview, in Part II, these same features of the

---

81 Being object-oriented and non-situational, imaginativist ideal conditions contrast with other proposals for ideal conditions for dispositional theories of value. Specifically, the imaginativist conditions are quite different from some idealizations that privilege rationality and full information. For example, the ideal conditions in the dispositional theory of value advanced by Michael Smith (1994, 158-161; 2002, 329-335) are not object-oriented but are situational. Smith spells out the idealization relative to a set of circumstances that the evaluator may find herself in. The full information relevant to one set of circumstances may be different from the full information relevant to another set of circumstances. However, given a set of circumstances, the idealization does not vary with respect to different possible objects of evaluation.
idealization will serve to distinguish desirability from other sorts of egocentric value, especially well-being.

I.3 Pluralism

Pluralism, in Berlin’s sense, requires at least this much: First, there are several, perhaps many, values. Second, these values may, in some important sense, be in conflict. I will now show how an imaginativist dispositional theory like the one just sketched explains these features of pluralism.

First, and most dramatically, the object-oriented idealization allows that two things may be values even if they cannot both be realized together. This was already implicit in the example above about the value of universal healthcare despite its costs. Of course, it is not impossible for universal healthcare to be less expensive. This might be the case if medical research and technology were not expensive, or simply if people were less frequently ill. Acquaintance with the proposition that there is universal healthcare does not require acquaintance with its costs. So, it is not terribly surprising that both universal healthcare and freedom from financial burden are values.

In other cases, both of two things may be values even though not even a drastic change in the way the world operates could make them compatible. To adapt another example from Berlin (2013, 12), it may be impossible for a system of criminal justice to be rigorous and exacting, on one hand, and merciful on the other. (Contrast the Code of Hammurabi, on one hand, with movements to replace retribution with rehabilitation, on the other.) This, in itself, does not entail that rigorous justice and mercy cannot both be values. Whether either or both are values would depend, of course, on whether full imaginative
acquaintance with (a suitably expatiated and refined specification of) each may cause us to value it.

Countenancing incompatible values comes from imaginativism’s object-oriented idealizations and its exclusion from those ideal states awareness of information about anything other than what is being evaluated. Again, what is full is the acquaintance, not the information. Unless they are specified as part of the putative value, costs and incompatibilities are ignored.

The plurality of incompatible values would not be very important if it had no bearing on the role of these values in our lives. However, it does indeed matter. An immediate consequence of the existence of incompatible values is that a person may be only minimally, or perhaps not at all, blameworthy for failing to pursue some particular value. If two genuine values are incompatible, then, ipso facto, they cannot both be realized. But if it is impossible to realize both, it may not be rational to pursue both. Of course, there may be cases where simultaneously pursuing two incompatible values improves the chances of getting one or the other. But at the point when this no longer holds, if a person chooses to forgo one or the other, we ought not blame her for this. Imaginativism yields, then, this welcome feature for a pluralist theory of value: Not all values are to be sought by everyone.

Just because there are incompatible values, it does not follow that we do not have any reason to prioritize one over another. If that is right, it may make the pluralism we have here a little less exciting. If we have something like a lexical ordering or some objective weighting of the values—something to tell us which were more valuable than which others—then we have a significant sort of unity after all. We could think of values as unified in
virtue of each being located within in a single organized hierarchy. To put the point another way: A view that does not posit significant incommensurability among values is, intuitively, not so pluralistic.\(^{82}\)

I will not argue here that there is no general way to rank or prioritize all the values encompassed by the imaginativist’s theory. Indeed, there might a be a moral basis for some such ranking.\(^{83}\) I simply want to point out that is difficult to find any basis for an overarching scheme of prioritization within the imaginativist dispositional theory of value itself. Seeing why a few of the seemingly promising possible bases are faulty helps us see in what sense the imaginativist dispositional theory yields a pluralistic conception of value.

The first thought we might have is that we could rank values according to how much we would value them under ideal conditions. So, if, under ideal conditions, we would value A more than B, then A would be more valuable than B. But this idea runs into difficulties when we ask what the relevant ideal conditions are supposed to be. Because of the object-orientation of imaginativist idealization, the ideal conditions for valuing A and the ideal conditions for valuing B will be different. So, neither of those sets of conditions can be the ideal conditions we are looking for, since neither set of conditions is likely to be neutral between the two values.\(^{84}\) We could try to identify some other set of conditions appropriate

---

\(^{82}\) Note, though, that we probably ought not hold that incommensurability of values is a necessary condition of pluralism. As Chang (1997, 14-17) argues, pluralism itself does not provide grounds for positing incommensurability. For present purposes, the important point will be that the imaginativist dispositional theory does entail one sort of incommensurability, and this bolsters the theory’s pluralistic credentials.

\(^{83}\) I understand moral value as a more narrow sort of value than what we are examining here.

\(^{84}\) If there were two values such that the state of full imaginative acquaintance with one were the same as full imaginative acquaintance with another, and if our dispositions to value or the strength of our valuing differed for the two values, then this would be a basis for ranking these two values. However, any such cases will be exceptional, and so this strategy would not provide a general way to compare any given pair of values.
for comparing our responses to the two values, but the imaginativist dispositional theory would not provide any grounds for privileging any other proposed “ideal” state.

Since we have found that comparing degrees of valuing from the standpoint of a single ideal state is unworkable, we might still hope to rank values according to how much we are disposed to value them. Instead of choosing a single ideal state for comparing several values, perhaps we could look at the strength of the valuing attitude elicited by the values in the conditions respectively ideal for each. This proposal—call it the strength of valuing proposal—suggests that, if under conditions of full imaginative acquaintance with V, we value V with strength X, under conditions of full imaginative acquaintance with W, we value W with strength Y, and X>Y, then V is more valuable than W. This proposal is more promising than the first, but I have a couple of worries about it.

One of my worries about the strength of valuing proposal comes from skepticism about the idea that the attitude of valuing (or any other pro-attitude) has a strength independently of its interaction with—e.g., overpowering, outweighing, preempting, trumping, etc.—other attitudes. The proposal seems to be that we discover the strength of the valuing elicited by one value, discover the strength of valuing elicited by another value, then compare those two strengths. But this seems to require that attitudes somehow have their degrees of strength intrinsically, not just in relation to other attitudes, and I do not see a good way to make sense of that. Perhaps, though, there is alternative way of working the strength of valuing proposal. The idea is still to start with the valuing of each object as elicited under its specific ideal conditions, but then we put the different appropriately elicited valuing attitudes into interaction with each other. Once removed from their respective ideal
circumstances and put together, we might find that one outweighs the other, and this gives us a way to measure strength of valuing. My problem with this is that I see no reason to equate the strength of an attitude formed in one set of circumstances with the strength of the same attitude later, once circumstances have changed. After all, as our situations change, our attitudes change. Unless attitudes have strength intrinsically, I do not see how to interpret the thought that an attitude formed in certain circumstances could keep the same strength when the circumstances change.

Another worry about the strength of valuing proposal is that, even if this procedure yielded degrees or some other measure of the strength of valuing, it is not clear that this information would tell us what we wanted to know. Specifically, it is not clear that the degree of valuing would measure the degree of value. After all, we are supposing that the ideal conditions, the conditions under which the valuing is to take place, differ. And some of these conditions might be more conducive to a strong valuing response than other conditions are. For instance, full acquaintance with a very complex proposition might require a degree of mental concentration that limited the person’s capacity for valuing. This limitation in the capacity for valuing would be because of the general damping of valuing capacity, due to intense concentration, not because of lack of sensitivity specifically to this particular value. In such a case, the strength with which we valued some proposition in the

---

85 I am also a little worried about how to identify an attitude over time. For the proposal in question to work, I think that the identity of an attitude across time would have to consist in more than just the conjunction of sameness of attitude type (whether desiring, liking, caring, etc.) and sameness of content (about this or that proposition).

86 If we had independent reason to hold that the attitude of valuing was always fairly stable in its strength, this proposal might be workable after all. Valuing, as understood in Svavarsdóttir (2014), might have this feature. Since I am aiming to be minimally committal about what sort of attitude or attitudes valuing is, I am in no position to rule this out.
particular conditions ideal for valuing it, may not be very indicative of the value of the proposition itself.

A third attempt to find in the imaginativist dispositional theory a basis for ranking values would be to use the theory to reach evaluations of propositions that themselves somehow expressed a ranking of the values. This is an interesting idea. It involves somehow representing a proto-comparison with a proposition and reaching the evaluative comparison by assessing the desirability of that proposition. Now, note that if we were trying to compare value A and value B, the proposition in question could not just be that \( A \) is more valuable than \( B \). In order to be fully imaginatively acquainted with this proposition we would already have to know what it was for one value to be more valuable than another; the account of evaluative comparison would be circular. But maybe there are other options for the proto-comparative proposition. Consider the proposition that we (say, humanity) prioritize fair distribution of primary goods over the creation of beautiful artifacts. It may be that, under conditions of full imaginative acquaintance with this proposition, we value it. If so, then that is a value, albeit a somewhat complex one. But the question is whether this tells us anything about the ranking of the values themselves. Does the value of us prioritizing A over B show that A is more valuable than B? I am not convinced that it does. Moreover, I have doubts about whether this strategy for comparing two values would generalize to yield a way of organizing all values into a hierarchy. That said, I do not have an argument that no approach along these lines could work. If some such approach were workable, we would be left with a mitigated form of value pluralism, a pluralism without incommensurability.
I.4 Intersubjectivity

Berlin emphasized pluralism and conflict among values, but he also took pains to emphasize that not just anything could be a value. Recall, he says, “the number of human values, of values which I can pursue while maintaining my human semblance, my human character, is finite” (2000, 12). A link between value and what it is to be human is, according to Berlin, what keeps value pluralism from being value relativism. On Berlin’s view, there is a single set of objective values. As I emphasized in the preceding section, the values in this set may conflict with one another. Furthermore, not everyone will value all of the values (nor could we, in all likelihood), and most of us will value some things which are not instances of the values. Still, the single set of values is not supposed to be relative to any person or group.

This looks like a problem. Or, at least it suggests a place where Berlin’s pluralism and an imaginativist dispositional theory might part ways. That is because the dispositional theory of value described by Lewis is inherently relativistic. The truth conditions of a statement about value depend on who the speaker is and who else is a member of the speaker’s group when she says ‘we’. But note, after all, a sense in which Berlin’s own view is itself relativistic: It is relative to humanity. Berlin’s view is not that the finite set of objective values is independent of any valuers. He does not think these values would be the same given any possible rational agents or any possible beings capable of valuing. Hence, if

---

87 Alas, Berlin does not say much about exactly what link he has in mind or, in general, what sort of link is sufficient to do this sort of work. In general, it is not at all clear to me that pluralism and relativism are opposing views in the way that Berlin seems to suppose. There is much more to say about the relationship between pluralism and relativism than I can say in this chapter.

88 Or, to put it another way, for any particular person, there is unlikely to be a perfect match between the values and that person’s values. See note 75.
by ‘we’ we mean all of humanity, this brings the imaginativist dispositional theory into alignment with Berlin’s pluralism. More precisely, if ‘we’ is all of humanity, then both the imaginativist dispositional theory and Berlin’s theory *are* relativistic, but the relativity is to the very same (large) group.

But should we say that the ‘we’ in the dispositional theory of value includes all humans? If it does, then something will not be a value unless every human being is disposed to value it under conditions of full imaginative acquaintance. Lewis is noncommittal, but he sounds a note of optimism, suggesting, “Maybe all mankind are alike” (1989, 126). Lewis recognizes the diversity in what people actually value, but his thought is that the differences in what we actually value could be due to people not valuing as they would under the relevant ideal states. In other words, it is not at all uncommon for people to value something that is not actually valuable. To this we can add the observation that the imaginativist dispositional theory is consistent with people valuing very different subsets of the full set of values, even though all of us would value each member of the full set under conditions of full imaginative acquaintance with it. Hence, it would not be incredible if we learned that there was enough psychological commonality in human dispositions to value to ground a fairly large, but finite, set of values.  

89 Since something being a value requires unanimity among those in the ‘we’, if we want a set of values of any decent size, including all of humanity in ‘we’ is almost certainly to include too much. Different dispositions to value are distributed unevenly across humanity. This, in itself, is not much of a worry. The thought would be that only those dispositions that are shared universally yield values, and there still may be plenty of these (though that is a decidedly empirical hypothesis). The bigger worry is that some particular human beings are pretty unusual. Even without considering the interference of evil demons or mad scientists, some beings who still count as human may not share all of the nearly universal human psychological dispositions. For example, traumatic experiences or irregularities in neural development could cause this. If that is so, then we should probably take ‘we’ to refer to *almost all of us humans*, or perhaps to *all human persons* (where the category of *persons* is defined such that being a member of that
Ultimately, how much diversity there is across humanity in our dispositions to value is an empirical question. If it turns out that we are extremely diverse in this way, this would have two consequences. First, unless we could identify some other aspect of our common humanity on which to ground values, it would mean that Berlin was wrong about there being a large set of values common to all humanity. Second, it would mean that letting ‘we’ refer to humankind in the dispositional theory, resulted in there being no, or very few, values. The problem for Berlin—since he is very much committed to there being common human values—would be more serious than the problem for the proponent of the dispositional theory. It is open to the proponent of dispositionalism to acquiesce to more relativism, allowing the extension of ‘we’ to vary with the speaker. Then dispositions would not have to be shared so widely in order to ground genuine (although more relativistic) values.

The intersubjective element of Berlin’s view is not limited just to the claim that there is a single set of human values. He emphasizes that the behavior of someone who embraces values different from those we embrace is intelligible to us. Understanding among humans is possible even when we are in the grips of very different sets of values (so long as all these “values” are indeed genuine values). The imaginativist dispositional theory is well-suited to explain this as well.

Suppose that Amy (a human person) values V, because V is a value. In other words, Amy values V because Amy (like all humans) is disposed to value V under conditions of full imaginative acquaintance with V, and Amy has been very well acquainted with V. But

---

category carries some normative weight). I doubt this would bother Berlin.
suppose that I do not value V, even though V is a value. It may be that I have never given much consideration to V, or that I used to value it, but other concerns crowded it out of my valuational economy. In that case, Amy’s orientation to V is not something with which I sympathize. However, I can come to sympathize insofar as I can approximate full imaginative acquaintance with V. If I can become fully acquainted (or nearly so) with V, my disposition to value V will kick in, and I will value it. Then my orientation to V will resemble Amy’s. I will value it the way she does, and I will understand her valuing by knowing first-personally what it is like. Thus, her valuing becomes intelligible to me at a deep level. Now, if and when I lose my full acquaintance with V (which, for all I have said, may happen promptly), I may no longer value V. Still, having once valued V in the way Amy does and remembering what that was like continues to help make her intelligible to me.

Now, this explanation of how someone’s valuing, which is different from my own, could become intelligible to me crucially required that I be able to achieve or at least approximate full imaginative acquaintance with the value in question. We stipulated that Amy had, at some point, achieved full acquaintance (or near enough) with V. Why think that I, or anyone else, would share this capacity with Amy? The key to the explanation is the fact that my state of full imaginative acquaintance with a particular value will be very much like anyone else’s state of full imaginative acquaintance with that value. In particular, our individual life circumstances—like our goals, our constraints, what count as costs and benefits for each of us, etc.—do not affect what would count as full imaginative acquaintance with a particular value. The ideal conditions have much more to do with the thing being
evaluated than with who is doing the evaluating. This, of course, is just to reiterate that the imaginativist idealization is object-oriented and non-situational. As a result, the idealizations, though there are many (one for every value), are common ground among us all, and states any of us may aspire to achieve. This, I think, helpfully illuminates one of Berlin’s remarks quoted earlier: “[I]f a man pursues one of these values, I, who do not, am able to understand why he pursues it or what it would be like, in his circumstances, for me to be induced to pursue it. Hence the possibility of human understanding” (2000, 12).

Part II: Desirability

II.1 From the valuable to the desirable

The notion of value discussed in Part I of this chapter was intended to be non-egocentric value. The value was not value for any particular person. In Part II of this chapter, my focus will be egocentric value. My aim here is exploratory and constructive. In Part I, I showed how an imaginativist dispositional theory could explain and elaborate some of the distinctive features of Berlin’s conception of non-egocentric value. If that is right, might the same materials yield a similar theory that defines a useful conception of egocentric

90 Now, granted, some feats of acquaintance may be easier for some of us than for others. It may be relatively easy for me to become fully acquainted with something that is similar to other things with which I am familiar. To become acquainted with something that is quite different from anything I have experienced or otherwise have been acquainted with may require some work. But we were never supposing that intersubjective intelligibility would be effortless.

91 To read this remark in a way that coheres especially well with the account of intersubjective intelligibility I have been laying out, we should take “in his circumstances” to imply circumstances ideal for valuing that value. This would make the person here much like the Amy character, whom I described as valuing V because V is a value. However, if the circumstances that induced the other person here to pursue the value in question were very particular to that person and unlike the circumstances ideal for valuing that value, then I am less likely to be able to understand why that other person values as he does. I think Berlin would appreciate this qualification.
value? The answer, I think, is yes. The goal of Part II is to show how a small change in the theory described in Part I yields a conception of desirability, according to which it is a type of egocentric value, distinct from both well-being and the goodness of a life. I will also draw on Part I’s discussions of pluralism and intersubjectivity to show what makes desirability a significant evaluative category. By the end of this chapter, we will have an important, though philosophically under-appreciated, notion of egocentric value clearly before us.

I locate desirability’s philosophical starting point in the work of Sidgwick. In Book I, Ch. IX of *The Methods of Ethics*, Sidgwick develops and considers several possible accounts of a *person’s own good*. These are conceptions of egocentric value. The notion which Sidgwick finally deems “authoritative” is a person’s *good on the whole*. Roughly, a person’s good on the whole is what the person would desire for herself if she were fully rational and fully informed about her life and the consequences of different possible actions (1907, 111-112). This is a dispositional theory of a particular sort of egocentric value. Notably, in this theory of a person’s good on the whole, the ideal conditions involve rationality and information, not imaginative acquaintance. But before laying out this account of a person’s good on the whole, Sidgwick presents a dispositional account of a different notion of egocentric value, and this account bears a striking similarity to Lewis’s dispositional theory of value, which we have been scrutinizing. Sidgwick talks about this kind of value as what is *desirable* for a person.92

---

92 If a person’s good on the whole is different from desirability, then might it be a conception of one of the other kinds of egocentric value I have discussed in this dissertation, i.e, well-being or life-goodness? I think we should say that a person’s good on the whole, in Sidgwick’s sense, is what would actually increase her life goodness. This is not to say that a person’s good on the whole stands in a general contribution
It is unclear exactly what role Sidgwick intends his account of desirability to play in his overall exposition. On one reading, perhaps the most straightforward, desirability is a notion of egocentric value that Sidgwick considers and rejects in favor of his eventual definition of a person’s good on the whole. Regardless, as we shall see, desirability is an important evaluative category in its own right.

Sidgwick characterizes the desirable as “what would be desired, with strength proportioned to the degree of desirability, if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition” (1907, 110-111). Clearly, this is a dispositional account of desirability. It says that what is desirable for a person is what she is disposed to desire, what she would desire in a suitably ideal state. I put aside the issue about what attitude is involved. Sidgwick mentions the attitude of desiring, but I will continue to operate with the attitude of valuing (carrying over the qualifications and hedging from section I.2 above). Here, as in Part I, my focus will be the idealization. What sort of ideal state does Sidgwick have in mind? The first thing worth pointing out is that the idealization does not seem to be imaginativist, at least not exactly. The talk of “a perfect forecast” of the realization of the desirable thing suggests that the ideal state includes full (or, at least, enhanced) information about the putative value. On the other hand, that perfect forecast is described as “emotional as well as intellectual” might suggest that to “possess” a perfect forecast is to be in a state of heightened awareness of or acquaintance with what is being evaluated.
Though it is unclear to what extent Sidgwick's account is imaginativist, it is clear that the account shares the two notable characteristics of the imaginativist dispositional theory described in Part I. First, the ideal state is object-oriented; it is specified schematically. So the ideal state will vary with the object being evaluated and can be specified completely only in terms of the person’s relationship to that object.

Second, the ideal state is non-situational. We can see this in the condition that the object be “judged attainable by voluntary action.” Here Sidgwick recognizes that our dispositions to desire are sensitive to our beliefs about what is and is not attainable. But, according to the way he is thinking about desirability, whether a person is actually in a position to attain something is irrelevant to whether it is desirable for her. Hence, the ideal conditions factor out the issue of whether the object is judged attainable by stipulating that it is. This has the effect of abstracting away from the present practical circumstances of the person. Furthermore, not only are the preconditions of attainment ignored; so are the consequences of it. This comes out when Sidgwick moves on to characterize his second notion of egocentric value, the notion of a person’s good on the whole. He prefaces the characterization of that second notion by distinguishing it from desirability. He notes that even if something is desirable, and is indeed such that its realization were just as favorable as it was forecast to be, its realization may have “concomitants and consequences” that are unfavorable for the agent (1907, 111). A person’s good on the whole depends on these factors, but desirability is independent of them. Hence, we should understand the ideal conditions in the dispositional theory of desirability to bar any acquaintance with any contingent consequences of the realization of what is being evaluated.
I will not examine the details of Sidgwick’s account of desirability any more closely.\textsuperscript{93} Instead, my aim has been to show how a particular sort of dispositional theory—one that involves an object-oriented and non-situational idealization—may point the way to a distinctive sort of egocentric value. Furthermore, though I will not pursue this exegetical point, Sidgwick’s distinction between desirability and a person’s good on the whole prefigures the distinctions I wish to draw between desirability and other kinds of egocentric value.

Putting aside Sidgwick now, we can get at desirability another way by changing one element of the Lewisian imaginativist dispositional theory of value discussed in Part I. So far I have given scant attention to the question of what category of things might be valuable. Lewis raises this question, but he does not suggest, as I will, that our answer to this question may lead us to either an egocentric or a non-egocentric conception of value.

In Part I, I assumed that the category of valuable things was \textit{propositions}. So, for instance, when I talk about the valuing sheltering refugees, more precisely what I value is \textit{that refugees are sheltered}. But Lewis (1989, 118) points out that this is not the only possibility for the type of thing that I might value, and so it is not the only option for the type of thing that might be a value. The content of some of my valuing has an ineliminable egocentric element. I might value not just that the world be a certain way, but that \textit{I} be a certain way. The thought is that there is a difference (though it may not always be a

\textsuperscript{93} One other element of Sidgwick’s definition of desirability is the bit about the proportionality of the degree of desire under ideal conditions to the degree of desirability. I part ways with Sidgwick here. First, I suspect a ratio scale, as Sidgwick seems to have in mind, is way too much to ask for in the measurement of desiring (or valuing). The discussion about incommensurability in Section I.3 above applies here, too. In general, regarding desirability I accept what Lewis says of value: “Our present business is not with the balancing, but with the prior question of what values there are to balance” (1989, 124).
difference that makes a practical difference) between, say, me valuing that Owen be courageous and me valuing that I be courageous. My proposal is that when we restrict the imaginativist dispositional theory of value, the theory discussed in Part I, to just our dispositions to value essentially egocentric contents, this opens a path to a theory of egocentric value in the vein of Sidgwick’s notion of desirability.

II.2 Egocentric value: relational, not relative

Return now to the original statement of Lewis’s dispositional theory of value from the beginning of section I.2 of this chapter. It was as follows:

Something of the appropriate category is a value iff we would be disposed, under ideal conditions, to value it.

Compare two different things that might be values, one of which we might express as “that people take walks”, and the other as “taking walks”. The first possible value is simply an ordinary proposition. It is indeed a value iff, we (i.e., myself and everyone my group, which I continue to assume includes almost all humans\(^\text{94}\)), under conditions of full imaginative acquaintance with people taking walks, value that. Note that, strictly speaking, this does not especially require us to value going on walks ourselves; it does not require any of us to value taking walks ourselves more than we value anyone else taking walks. The second possible value, going for walks, can be understood as a de se proposition, a proposition of the sort most naturally expressed with a statement involving a first-person pronoun.\(^\text{95}\) This purported value is indeed a value iff, each of us, under conditions of full imaginative

---

94 For a qualification, see note 89.
95 For discussion, see Lewis (1989, 118-121).
acquaintance with the *de se* proposition *I take walks*, would value that. According to this picture, which I take to be the one Lewis intends, regardless of whether the thing being evaluated is an ordinary proposition or an egocentric proposition, the relativity is the same. The truth of a statement about value is relative to the speaker and those included in the speaker’s group.

However, we can produce a different sort of theory if we put special emphasis on the essentially egocentric contents of *de se* propositions. In one sense, if we both value taking walks, we value the same thing. But note that, in another sense, we do not. If I value taking walks, then what I value essentially involves me, and if you value taking walks what you value essentially involves you. It is not so much that we value the *same thing*, as that we value *alike*. When we emphasize this way in which the contents of the *de se* thoughts of one person differ from the contents of the *de se* thoughts of another, we see a sense in which there is only one person who could value, or even entertain, the thought I express when I say, “I go for a walk,” and that person is *me*. If there is a thought that only I can entertain, then the *we*, the group of people who might be disposed to value it, collapses to just *I*. Along these lines, my proposal is that if we start with Lewis’s definition of value and then restrict the category of valuable things to essentially egocentric contents understood this way, that yields this definition of desirability:

Something is desirable (for a particular person) iff the person would be disposed, under ideal conditions, to value it.

---

96 Lewis talks about people who desire the same *de se* proposition as desiring alike (1989, 119).
This definition of desirability is almost just like the definition of value in Part I. The ideal conditions and the attitude of valuing remain the same. The principal difference is not revealed in the definition itself: The something here is to be understood as something like a proposition but with essentially egocentric content. The more apparent difference, that this definition is indexed to an individual, follows from this principal difference, and I will return to this point shortly. Before that, it will be helpful to expand on the idea of essentially egocentric content. As John Perry (1979), famously pointed out, some of our thoughts include an essential egocentric element. In my own mental economy thoughts about Owen and thoughts about me may sometimes play different roles. In a dramatic case, like if I were to forget who I am, the thought, for instance, that the murderer is outside Owen’s door may not move me appropriately when I hear a knock. Furthermore, even if I have not forgotten who I am, thinking of myself in the third person can serve to detach my thinking from my practical and emotional dispositions.

Along the same lines, an egocentric desire, a desire essentially about myself, may differ from a desire about an ordinary proposition regarding the same person. I could desire that the panel discussion go well, and hence I might desire that both Janet’s presentation is interesting and that Owen’s presentation is interesting. In this way, I desire something

---

97 Given the similarity with the definition in Part I and the fact that the definition invokes valuing, not desiring, why call this value it “desirability” at all? Well, first of all, I need a term to distinguish this egocentric value from the non-egocentric value in Part I. I might just call this notion “egocentric value”, but I want to reserve that term for a general category that includes desirability, well-being, and life goodness. So, we need a different term. ‘Desirability’ is suitable, and we find in Sidgwick a precedent for using the term this way.

98 Unless, of course, the valuing of egocentric contents is a different type of attitude than the valuing of non-egocentric contents.

about myself *qua* a part of the world. But I might also care about giving an interesting presentation, and so desire something about myself *qua* myself. What goes for desire here, goes also for valuing and other pro-attitudes.

Now we can see the difference in the respective definitions of desirability and value. When I say, “It’s desirable that my presentation is interesting,” this is true iff I would be disposed, under conditions of full imaginative acquaintance with my presentation being interesting, to value my presentation being interesting. But if I say, “It’s valuable that Owen’s presentation is interesting,” then this is true iff each of *us*, i.e., all humans or all human persons,100 would be disposed, under conditions of full imaginative acquaintance with Owen’s presentation being interesting, to value Owen’s presentation being interesting. 101

Now it becomes clear why the egocentric content in the definition of desirability requires mention of a particular individual on the right-hand side of the definition. When I say, “It’s desirable that my presentation is interesting,” this can be about only one person, viz., Owen. When another person utters the same sentence, it is about that other person. Furthermore, there is only one person who can value egocentrically that Owen’s presentation is interesting, viz., Owen. Any other person’s perspective on Owen’s presentation would lose the essentially egocentric element. Hence, since the content that is being assessed is essentially egocentric, we can see why the right-hand side of the definition of desirability has to mention a particular individual.

100 Again, see the caveat at note 89.
101 Now, of course, in ordinary English, ‘valuable’ and ‘desirable’ are used more or less interchangeably. So, whether a particular utterance is understood as being about value or desirability will depend on what was the purported value and whether it was essentially egocentric.
But that does not yet prove that the value in question is egocentric value, in the sense that a person’s well-being and the goodness of her life are kinds of value for her. We have not yet seen why satisfying the right-hand side of the definition would entail value for the person mentioned. Perhaps the mention of a particular individual on the right-hand side of the definition is just an indicator of some kind of speaker relativism. The definition of desirability could be relativistic, relative to the speaker, in the same way that the theory of value in Part I was. For desirability, we want a theory according to which the truth of an evaluation is relative to the person whose value the statement is about. To put it another way, instead of a relativism, we want a theory of relational value. My claim here is that, once we are dealing with egocentric content, we get egocentric value, i.e., relational value. To see this, we can observe how the definition of desirability works when people try to make evaluations together.

First, review once more how evaluations work according to the earlier, non-egocentric definition of value. Suppose I say, “Playing music is valuable.” Now, since we are understanding the objects being evaluated as propositions, this means the same as my statement, “It is valuable that people play music.” Now, if we follow the original dispositional analysis, we will say that my statement was true iff were we (I and everyone in my group) fully acquainted with it being the case that people play music, we would each value that. Now, suppose you wish to agree that my evaluation was correct, and you also utter the statement, “It is valuable that people play music.” Then your statement is true iff

102 Though, again, the relativity is mitigated to the point of invisibility by a sufficiently expansive reading of ‘we’.  
were you (you and everyone in your group) fully acquainted with it being the case that people play music, you would each value that. The analysis is clearly speaker-relativistic. However, if our groups are the same (which, of course, they would be if humankind is the one relevant group), then the relativism is invisible, and my statement was true just in case yours was.

What if my original statement had been, “It is valuable that Owen plays music”? Then the analysis will go just the same. My statement is true iff I am (or my group is) disposed to value Owen playing music. And if you utter the same statement, what you said is true iff you are (or your group is) disposed to value Owen playing music. Again, the analysis is relativistic, but if our groups are the same, then my statement is true when yours is. So far, this has been just to apply the original non-egocentric theory.

Now, suppose I make an egocentric evaluation, saying, “Playing music is desirable.” This would mean the same as my statement, “It is desirable that I play music.” According to the definition of desirability I am proposing, the statement is true iff were I fully acquainted with myself playing music, I would value that. Here is the issue: Suppose you want to agree here that, yes, that would be desirable for me, and so you say, “It is desirable for Owen that he play music.” The crucial question is about how we are to interpret your statement. Whose disposition to value matters here?

Now, if we were dealing with the relativism of the original theory, we would say that the truth of your statement depends on your (and your group’s) dispositions to value. But that does not make sense here. Suppose you were to think you could test your statement by the same method that I test my statement about myself, and so you try to see whether it is
the case that, were you fully acquainted with yourself playing music, you would value that. Clearly, whatever you come up with will be irrelevant to my statement. That is because the content of your imaginative acquaintance was wrong. You made your thoughts about you, instead of about me. But your evaluative statement was supposed to be about me.

According to another possible interpretation, still of the relativist variety, you should see whether you were disposed such that, were you to be fully acquainted with me playing music, you would value that. But this does not work either. Your evaluation is about the right thing, in one sense, since this time you were thinking about me playing music. But, in another sense, your evaluation was not about the right thing; the perspectival element of the content is incorrect. You have lost the essential egocentricity of what was to be evaluated. You have simply reverted to the original non-egocentric dispositional theory of value and applied that to the proposition that Owen plays music. Thus, this interpretation simply ignores what is distinctive about the proposed definition of desirability.

We have found that neither of the two possible relativistic approaches yields an acceptable interpretation of a person’s statements about what is desirable for another. So, instead, we should conclude that the correct interpretation of your statement about what is desirable for me will not involve your dispositions at all; it will be about mine. Your statement, “It is desirable for Owen that he play music,” is true iff were Owen fully acquainted with himself playing music he would desire (of himself) that he play music. Hence, though the original, non-egocentric dispositional theory of value was relativistic, the dispositional theory that ranges over essentially egocentric contents is relational. Therefore,
according to the definition I am proposing, desirability is relational value, value for some person.

At this point, even if we accept that desirability is relational, we might worry about whether it is really relational value. When something is desirable for a person, we might wonder, is that really a sense in which it is good for her? I think that, if dispositional theories of value have anything at all going for them, then it must be. Now, of course, I have not given an argument for dispositionalism. Rather I have been using a dispositional theory developed in one area to shed light on another area. But if the core thought behind dispositionalism is right—if it is right that what is valuable has to do with how we are disposed under ideal conditions to value—then, when the valuing is essentially self-regarding, it is hard to see how this would not be value for oneself. Some hesitation here might be due to not finding the dispositional approach compelling. However, I suspect that most hesitation on this point is due to a (mistaken, I believe) assimilation of all egocentric value to well-being. It is to this issue I will turn next.

This concludes the promised introduction and definition of desirability. What remains is drawing out implications of conceiving of desirability this way.

II.3 Egocentric value need not be about well-being.

As I discussed back at the beginning of Chapter 2, there is a common tendency to think that someone’s well-being during her life and the goodness of her life are basically the

104 This, I grant, is a big if. There are critics of dispositionalism who doubt that what we would do under some ideal circumstances has any direct bearing on what is valuable or on what we should do in actual circumstances. See, e.g., Hubin (1996) and Enoch (2005). An attempt, not just to illuminate desirability, but to defend this dispositional account of it, would require responding to this general and important worry.
same. I argued at length that this is a mistake. In some cases, the mistake is an instance of the more general mistake of thinking that all egocentric value is the same. Committing this general mistake would lead one to think that evaluations of what is desirable for someone are ultimately about her well-being or the goodness of her life. I believe that this general mistake is a common one. My grounds for believing this are autobiographical. When I first began trying to understand what Sidgwick’s notion of desirability might amount to, people with whom I discussed this would say things like, “Oh, you mean pro tanto well-being.” Although the notion of pro tanto, as invoked in moral philosophy, is most commonly applied to reasons, it is not hard to imagine how the pro tanto operator could apply to well-being. The idea would be that something advances my well-being, pro tanto, just in case realizing that thing, all else equal, increases my well-being. That is, we are capturing the significance of “pro tanto” by way of a ceteris paribus qualification. This leads naturally to the thought that we are to understand desirability in terms of well-being by way of a value contribution principle (VCP) of the sort introduced in Chapter 1.

The question, then, is whether desirability is related to either well-being or life goodness by a VCP. Upon examination, the clear answer will be that it is not. To push our intuitions in that direction, consider an example. Suppose that we are wondering whether it is desirable for me to drink a beer. More precisely, we are evaluating the desirability for me of the thought that I am drinking a beer. Note that each of these statements seems totally sensible:

Drinking a beer is desirable, but I have had too many already.
Drinking a beer is desirable, but I am walking on a treadmill.

Drinking a beer is desirable, but I am in the middle of a long drive.

Now, in contrast, notice that each of the following statements (assuming we read “all else equal” in the strict sense I have been using throughout this dissertation) is somewhat confusing:

All else equal, I would be faring better if I were drinking a beer, but I have had too many already.

All else equal, I would be faring better if I were drinking a beer, but I am walking on a treadmill.

All else equal, I would be faring better if I were drinking a beer, but I am in the middle of a long drive.

My argument here will not ultimately rely on our intuitions about statements like these. However, if your intuitions match mine, there is something confusing about the second set of statements, but not the first. We now have the resources to explain this. The idea is that the basis for an evaluation of desirability (that is, what figures into full imaginative acquaintance) is non-situational, not sensitive to features of one’s practical situation, whereas what affects a person’s well-being at any given time may be very situational. Hence, each of the statements in the second set, in effect, says something like this: *All else equal, I would be faring better in this situation if I were drinking a beer, but this situation is not one in which it would be good for me to drink a beer.* If the “all else equal” is
read in the strict sense of Chapter 1, this verges on incoherence. I find this to be fairly persuasive evidence that desirability and well-being are distinct evaluative categories, but I grant that intuitions may vary considerably here. So, I will proceed to the main anti-VCP argument.

If what is desirable for me stood in a contribution relationship with my well-being or the goodness of my life, then any particular desirable thing would also stand in such a relationship. Now, speaking personally, one thing that is very desirable for me is going for walks. That is, if I achieve a state of full and vivid imaginative acquaintance with the thought that I go for a walk, I value this. So, if desirability stood in a VCP relationship to well-being or life goodness, we would expect at least one or more of the following principles to be true:

(GW-CP) During some period of time, if during that time I am going for a walk, then I have a higher level of well-being during that period than during a corresponding possible period that is exactly similar except that I am not going for a walk.

(GA-CP) In a life, if during some period in that life I am going for a walk, then that life has a greater amount of aggregate well-being than a possible life that is exactly similar except that during the corresponding period I am not going for a walk.

(GL-CP) In a life, if during some period in that life I am going for a walk, then that life is a better life overall than another possible life that is exactly similar except that during the corresponding period I am not going for a walk.

Each of these three putative VCPs is false. It is not hard to come up with counterexamples. Here is a scenario that is likely (once all the details are filled in) sufficient
to falsify all three. Suppose that immediately before the period in question I just returned from a very long walk. If so, then taking another walk immediately is unlikely to increase my well-being or add anything to my life, even if all else is equal.

Thus, the argument that desirability is not just what, *ceteris paribus*, increases well-being or the goodness of a life was quite simple. However, it was not an argument we were in a position to make at the beginning of this chapter. At that point, any purported counterexample would have been too contestable. The tendency to assimilate all egocentric value to well-being or life goodness is very strong. And so, any example that seemed to show that something desirable for a person did not, *ceteris paribus*, increase her well-being might have been taken just as evidence that the thing was not desirable for her after all.

Now, undoubtedly, the things that are desirable for me, such as going for walks, often *are* important for my well-being and *do* bear on the goodness of my life. However, what the counterexample just sketched shows is that the relations between desirability and well-being and between desirability and life goodness are not as straightforward as the contribution relation. Furthermore, given the relative simplicity and mundanity the present counterexample, compared to the complexity and peculiarity of the counterexamples to WL-CP in Chapter 2, it seems that the connections between desirability and well-being and between desirability and life goodness are much less direct and more tenuous than whatever connection holds between well-being and life goodness. This suggests that our strategy for investigating desirability should be to focus first on desirability itself, instead of on its relations to other sorts of egocentric value. Once we better understand the role desirability
plays in our lives, we will be in a better position to figure out how the different categories of egocentric value are interrelated.

Recognizing that desirability is not related to well-being or life goodness by a VCP serves a propaedeutic, ground-clearing purpose for an unprejudiced examination of desirability. If we are indeed to get a handle on the role and significance of desirability, then we should be careful not to blind ourselves to various theoretical possibilities in advance by thinking of desirability merely in terms of contributions to well-being or life goodness.

II.4 Parallels between desirability and (non-egocentric) value

If I am right, there is a lot to say about desirability, its importance, and how it relates to other evaluative categories. Most of this will have to await future work. At the present, I want to see what light my earlier discussion of the imaginativist dispositional theory of value may shed on desirability. With respect to the earlier account of value, I discussed pluralism and intersubjectivity. I will briefly revisit each of these to see how they look in the context of desirability.

First is pluralism. The imaginativist dispositional theory of value yielded a pluralistic conception of value because of its object-oriented specification of the ideal state. All of the relevant arguments—since they did not rely on an exclusion of egocentric content from the things being evaluated—transfer from the case of non-egocentric value to the case of desirability without alteration. So, one consequence is that, as in the case of non-egocentric value, the imaginativist idealization makes it possible that both of two things could be desirable for me, despite the fact that it would be impossible for me to have both. To take a quite mundane example, having a healthy, kempt front lawn is desirable for me.
However, it is also desirable for me that I have no lawn whatsoever. Clearly these are not co-realizable. But I do find that when I imagine, as vividly as I can muster, each of these possibilities, respectively, I desire each of them. Furthermore, for the reasons stated in Section I.3, from the standpoint of desirability alone, these values are difficult to compare, perhaps incommensurable. Of course, that is not to say that they would be equally good for my well-being. Furthermore, they may not be equally choice-worthy options for me. When situated within the practical context of my actual life, extrinsic costs and benefits bear on the situation in a way that makes one of the options clearly preferable. There are many similar examples. For instance, living in a rural area is desirable for me, and so is living in an urban area, though living in a suburb is not desirable for me.

Such examples are reflective of how desirability appears in our lives. For any given choice I might make, I may have several desirable options. Viewed individually and in isolation, each option would be welcome if realized. After all, full acquaintance with it would lead me to value it. But viewed in the larger context of my life, some of the desirable options will make more sense than others. As I blaze a trail through life, I hope to avoid what is not desirable, and I try to realize, achieve, experience, or acquire, some of the things that are desirable. Which of these I choose may depend upon what more general things are desirable for me, and also which non-egocentric values I embrace. For instance, if a fast-paced social life is desirable for me, or if I recognize the value of cities as a way or organizing human interaction, either of these might make it more likely that I would opt for no lawn instead of a kempt lawn and urban living instead of rural living. Pluralism in desirability means that there is a large variety of sensible options for me—too many and with too much
mutual incompatibility for me to attain, or even appreciate, all of them. As I travel through life, I acquire a particular collection of some of the many things desirable for me, and this gives shape and meaning to my life.

Moving on now to the issue of intersubjectivity, we find that desirability plays an important role here as well. First, we should note that desirability cannot provide quite as stable a basis for intersubjectivity as non-egocentric value can. According to the theory we examined, if something is a genuine (non-egocentric) value, then you and I are both disposed to value it. And this gives us a way, at least potentially, to achieve mutual understanding through sameness of valuing, since if both of us were to achieve full imaginative acquaintance with the value, both of us would value it. But this does not apply to desirability. Desirability is value for a particular individual; it is about her dispositions to value, not anyone else’s. So, for something to be desirable for me, it is not necessary that anyone else be disposed to value it (nor could they be, on account of the egocentric character of what is desirable).

Despite this limitation of desirability as a basis for intersubjectivity, it still has a role to play. Now, again, for reasons discussed in Section II.2 above, unlike with non-egocentric value, we cannot relate to others in terms of exactly what is desirable for each of us. After all, what is desirable for you is something essentially about you, and what is desirable for me is something essentially having to do with me. In principle, it cannot be the case that the same things are desirable both of us. Nonetheless, desirability is still a basis on which we might become more intelligible to each other. If I am pursuing something that is indeed desirable for me, then I am pursuing something that I can appreciate independently of
anything particular to my personal practical circumstances. That is because the ideal conditions in the dispositional theory are non-situational. So, if I am pursuing something that is desirable for me, I am pursuing something you could get yourself into a position to appreciate, too. Whether not you would, in that position, appreciate it depends on your particular dispositions. But you do not have to get all the way into my shoes to find out if what is desirable for me is desirable for you. Even if our life circumstances are so different that you cannot put yourself in my shoes, there are, as it were, other common shoes (i.e., the non-situational ideal conditions) that fit us both.

In this way, desirability provides a basis for mutual understanding that well-being cannot provide. In many cases, knowing what will increase your well-being depends on knowing the present practical context of your life. What you pursue to improve your well-being may depend on the projects in which you are engaged and on how your current constraints and incentives are structured. Hence, those of your pursuits that are motivated by a concern for your own well-being may be hard for me to grok without finding out all these things about you.

Returning to the examples from the beginning, I do think that going on walks, gardening, and reading good philosophy are each desirable for me. To reach this judgment, I did not need to think through how these things might stand in relation to other possible goings-on in my life. Just by being well-acquainted with them, I can see that they are desirable for me. And, furthermore, from your standpoint, when I value these things, for you to see what moves me, you do not need to know the details of my life circumstances. You just have to acquaint yourself with analogous versions of these things (where it is you
walking, gardening, and reading, instead of me) and have a psychology similar enough to mine that your dispositions to value are similar to mine.
References


