Reviving Hedonism

by

James Fanciullo

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2020 by the
Graduate Supervisory Committee:

Douglas Portmore, Chair
Cheshire Calhoun
Tyler DesRoches

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

According to hedonism about happiness, all and only enjoyable experiences are the basic constituents of one's happiness, and these experiences contribute to one's happiness just to the extent that they have a greater intensity or duration. After defending this view, I show that it must be amended to count as an equally plausible theory of what constitutes one's well-being. I then present two such amended versions of hedonism about well-being. The first, which I call objective hedonism, adds the claim that the objective worth of the things one enjoys also makes a difference to the extent to which an enjoyable experience contributes to one's well-being. The second, which I call reliabilist hedonism, adds the claim that one's evaluative intuitions about which things are good for one track which things have proven themselves to one to reliably lead to enjoyable experience. I conclude that reliabilist hedonism constitutes a revival of hedonism about well-being.

ACKNOWLEDGEMENTS

I owe the bulk of my gratitude to my committee members, Cheshire Calhoun, Tyler DesRoches, and—most of all—Doug Portmore, who also advised me in writing this dissertation, for very many helpful comments and discussions. I am also extremely grateful to Richard Kraut, not only for many helpful comments and discussions, but for advising me during my two year stay at Northwestern University, during which this dissertation was written. I additionally owe thanks to audiences at Lingnan University and the National University of Singapore for helpful thoughts and discussions.

Regarding my intellectual debts, my thinking in this dissertation was stimulated by the work of many contributors to the study of happiness and well-being—far more than I can try to list here. Most notably, I was inspired and aided by the work of Roger Crisp, Fred Feldman, Chris Heathwood, and Eden Lin. Perhaps most of all, I benefitted from Richard Kraut, who happened not only to publish an innovative book on well-being, but to teach a class on "hedonism and its critics," as I just began writing this dissertation. The influence of both these happy coincidences on the dissertation cannot be understated.

Finally, for their undying love and support, I thank my family, especially my siblings Christina and Matthew Fanciullo, as well as my closest confidants, Dylan August, John Evans, Andrew Faragon, Nick Flaherty, Connor Gemerek, Conor Hurley, Matt Jones, Dan Lonky, Nick Muia, and Matt Shamlian. This dissertation surely would not have been completed without them, and I wish for them lives full of happiness of the most reliable sort.

Chapter 9 is drawn from my 2020 paper, “Do Fitting Emotions Tell Us Anything About Well-Being?” *Utilitas* 32: 118-125. It appears here by kind permission of Cambridge University Press.

TABLE OF CONTENTS

CHAPTER Page

PART I

1 INTRODUCTION 2

1.1 Three Taxonomies of Theories of Well-Being 3

1.2 A Prevalence-Based Classification: The Usual Suspects 8

2 ENJOYING THE GOOD 20

2.1 Three (Failed) Formulations 23

2.2 Two Additional Formulations 29

3 PLEASURE, DESIRE, AND HAPPINESS 42

3.1 Does Happiness Precisely Track Welfare? 50

3.2 The Nature of Attitudinal Pleasure 59

3.3 The Belief-Desire Theory of Happiness 65

3.4 Some Unsuccessful Objections 69

3.5 A Worry 73

3.6 The Genuine Desire Theory of Happiness 75

3.7 Taking Stock 76

4 THE GENUINE DESIRE THEORY OF WELFARE 78

5 SOME OBJECTIONS 98

5.1 Joke Book 98

5.2 Adaptive Preferences 102

5.3 What About the Animals? 105

5.4 The Euthyphro Objection 108

5.5 Paradoxical Desires 111

CHAPTER Page

5.6 False Beliefs and Defective Desires: An Asymmetry? 113

5.7 Wrapping Up 115

6 WRAPPING UP AND LOOKING FORWARD....116

6.1 Motivating the Alternative....118

PART II

7 WELL-BEING AS RELIABLE HAPPINESS 123

7.1 A Hedonistic Metaethic 130

7.2 Hedonism and the Limits of Well-Being 140

7.3 Wrapping Up 147

7.4 Reliabilist Hedonism vs. the Genuine Desire Theory of Welfare....149

7.5 Looking Forward 153

8 HEDONISM REVIVED 155

8.1 Reliabilist Hedonism 159

8.2 The Objectivity Objection....162

8.3 The Externality Objection....169

8.4 The Enjoyment Problem 175

8.5 Conclusion 178

9 DO FITTING EMOTIONS TELL US ANYTHING ABOUT WELL-BEING?....180

9.1 Fuchs's Test and Its Significance....182

REFERENCES....194

Part I

1 Introduction

This is a dissertation about what we sometimes call *well-being* or *welfare*. These are the concepts we invoke any time we make claims about things being *in a person's interest*, being *prudentially valuable*, or being *non-instrumentally good for* the person. When we make these claims, we are trying to pick out the features of a life in virtue of which that life would be good, in itself, for the person living it. We are trying to pick out the basic constituents of a person's well-being.

Though we often make claims about the well-being of ourselves and others—about how we might improve it, why it has diminished, and the like—we seem to lack anything near a complete understanding of the nature of this concept. To be sure, we have a firm enough grasp of it to confidently claim that a person's painful, needless suffering is bad for her; that it would be good, in itself, for her, were her suffering to be relieved. But *in virtue of what* would the relief of her suffering be good for her? What is it about this sort of suffering that *makes* it bad for her—or, perhaps, for anyone experiencing it? Without answers to these questions, answers to our everyday, practical questions—such as how we might improve a person's welfare, or why it has diminished—may be more difficult to provide. Yet, on these more fundamental questions, our pre-theoretical notion of well-being remains silent.

To answer these more fundamental questions, then, we need a more substantive theory of well-being. My aim here is to propose and defend two such theories. To anticipate: one of these theories claims that well-being consists in enjoyable experience, though the extent to which an enjoyable experience contributes to one's well-being partly depends on

the objective worth of the enjoyment's object; the second of these theories also claims that well-being consists in enjoyable experience, but rather than taking the objects of one's enjoyment to be relevant to one's well-being, it claims that our evaluative intuitions about well-being track which things have proven themselves to us to reliably lead to enjoyment. While the theories will reach similar conclusions about the goodness of any particular life, only the second, I will claim, does so in a way that is entirely faithful to hedonism's fundamental principle, on which enjoyment is the good.

Before saying anything else about these views, I should first explain the context in which they will be introduced. This will involve, not just a discussion of competing theories—and, if my own theories are to be motivated, of where these other theories fall short—but also providing some of the conceptual resources needed to mark the differences between the theories. I will begin, accordingly, by discussing a few useful ways of classifying theories, and then move on to introduce the competition.

1.1 Three taxonomies of theories of well-being

Theories of well-being essentially seek to explain both *which* things are non-instrumentally good or bad for a subject, as well as what *makes* those things non-instrumentally good or bad for the subject. That is, they seek to both *enumerate* the things that have non-instrumental value, and *explain* what it is about those things in virtue of which they have that value.[1]

[1] This distinction is introduced by Roger Crisp (2006, pp. 102-103). Contrary to my suggestion here, some people have claimed that certain theories of welfare—specifically, hedonist and objective list theories—are merely enumerative, while others—desire satisfactionist theories—are merely explanatory. See, for example, Fletcher 2013 and Woodard 2013. These people are mistaken, however, for reasons noted by Eden Lin (2017b). Properly conceived, every major theory of welfare is both explanatory and enumerative.

Even prior to specifying their content, these theories can be classified in terms of the *type* of enumerations and explanations they give. Some theories, for example, claim that there is only one *good-making property*—or one basic property in virtue of which a thing may be non-instrumentally valuable. Other theories claim that there is more than one such fundamental property. Theories of the former type are *monistic*; those of the latter type are *pluralistic*. To be a bit more precise: monistic views claim that there is only one type of thing that basically constitutes a subject's well-being; that there is only one type of thing that basically constitutes a subject's ill-being; and that anything that is non-instrumentally good (bad) for a subject is non-instrumentally good (bad) because it is a token of that type. Pluralistic views, on the other hand, claim either that there is more than one type of thing that basically constitutes a subject's well-being, or that there is more than one type of thing that basically constitutes a subject's ill-being, or both. So, for example, if a theory claims (implausibly) that well-being consists entirely in the possession of belief, this theory implies that a thing is a constituent of a subject's well-being if and only if and because that thing is a belief possessed by the subject. On this theory, there is just one good-making property—namely, the property of *being a belief*. The theory is, accordingly, monistic: anything that is prudentially valuable is prudentially valuable precisely because it is a belief. Alternatively, if we amended this view to claim that well-being consists in both the possession of belief and the possession of health, the theory would imply that a thing is a constituent of a person's well-being if and only if and because that thing is either a belief possessed by the person, or a healthy state of the person. The theory would then imply that there is more than one good-making property: a thing could be good because it is a belief, or it could be good because it is a healthy state. Since it proposes more than one good-making property, this theory is pluralistic.

We should be careful, however, not to confuse pluralistic views with theories of a different type: *hybrid* theories. Whereas pluralistic views claim that there is more than one good-making property, or more than one kind of thing that can constitute one's well-being, hybrid views typically claim that there is more than one condition that must be met in order for a single kind of thing to constitute one's well-being. If, for example, a theory claims that well-being consists in the possession of, not any beliefs, but just beliefs about one's own health, this theory implies that there is just one kind of thing that can constitute one's well-being—belief of the specified type—yet it also proposes more than one condition that a thing must meet to be a constituent of one's well-being: it must be a belief, and it must concern one's health. This is a monistic, hybrid theory. On the other hand, if a theory claims that a thing is a constituent of a subject's well-being if and only if and because that thing is either a belief about the subject's own health or a healthy state of the subject, the theory implies that there is more than one kind of thing that can constitute a subject's well-being—belief of the specified type, and healthy states—yet it also proposes more than one condition that one of these kinds of thing must meet in order to be a constituent of the subject's well-being. This is a pluralistic, hybrid theory. Thus we can see that whether a theory is monistic or pluralistic tells us nothing about whether it is hybrid or non-hybrid.

On the hybrid views just suggested, there are multiple conditions that are individually necessary and jointly sufficient for a single kind of thing to contribute to a subject's well-being. This is no doubt the most popular way of formulating hybrid theories. But not every hybrid theory must take this form. For it seems that hybrid views are instead characterized by their claim that the extent to which a thing contributes to a subject's well-being partly

depends upon whether, and to what extent, that thing has some further property.[2] It might be, for example, that a theory claims that well-being consists in the possession of belief, though the extent to which a given belief will contribute to one's well-being in part depends upon whether that belief is about one's own health. The theory might claim that belief of this latter sort is worth much more to your well-being than belief of other kinds, yet maintain that the other kinds of belief are nevertheless worth something to your well-being. Clearly, in this case, there is just one condition that a thing must meet to be a constituent of a subject's well-being: it must be a belief. Still, this is a hybrid theory, since the extent to which a belief will contribute to a subject's well-being partly depends on whether that belief has some further property: in particular, the property of being about the subject's own health. This way of characterizing hybrid views will come in handy later on.

A third and final way of classifying any theory of well-being is in terms of the role that theory assigns to a welfare subject's *attitudes* in determining the subject's level of well-being. While there is little agreement in the literature as to how exactly the distinction that I have in mind is to be formulated, we can understand it as follows.[3] On *subjectivist* theories of well-being, what is good for a subject is determined entirely by the subject's own attitudes. On *objectivist* theories, this is not the case. More precisely: subjectivist theories say that a thing is a constituent of a subject's well-being if and only if and because that thing is the object of one or more of the subject's actual or hypothetical attitudes. Objectivist theories, on the other hand, claim either that something other than the objects of a subject's actual or

[2] For a defense of this characterization, see Woodard 2015.

[3] See, for example, Sumner 1996 (pp. 38-41); and compare Keller 2009 (pp. 660-661) and Lin 2017a (pp. 354-355). My formulation follows the latter, more contemporary formulations of the distinction.

hypothetical attitudes basically constitutes a subject's well-being (that is, they reject the "only if" portion of subjectivism), or else that while such objects alone are the basic constituents of a subject's well-being, the relevant attitudes must also meet some independent evaluative standard (that is, they reject the "if" portion of subjectivism). Thus, for example, a theory on which well-being consists in the possession of healthy states will count as objectivist.[4] This is because the theory entirely ignores the subject's attitudes: any constituent of a subject's well-being must simply be a healthy state, which has nothing to do with the subject's attitudes towards those states. Similarly, a theory on which well-being consists in the possession of true beliefs counts as objectivist. For although this theory relies on the subject's attitudes (in particular, their beliefs) in determining the subject's level of well-being, it nevertheless implies that any constituent of a subject's well-being must meet the independent standard involved in being (not just a belief, but) a true belief. Alternatively, if a theory claims that well-being consists in the possession of whatever things a subject judges (or believes) god has willed for them, it plainly makes the subject's well-being entirely a matter of their attitudes: a thing is a constituent of their well-being if and only if it is the object of one of their beliefs about what god has willed for them. While each of these three theories have been monistic, and two have been non-hybrid, it should be clear that whether a theory is subjectivist or objectivist tells us nothing as to whether it is monistic or pluralistic, or hybrid or non-hybrid.

[4] This is the case, at least, so long as the criteria for a state's being healthy are to some extent independent of the subject's own attitudes towards the state. But no plausible theory of health would deny this.

1.2 A prevalence-based classification: the usual suspects

Though we have, for the purposes of illustration, to this point considered a number of substantive theories, none of these theories have seemed plausible in their own right. Turning our attention now to the specific content of theories of well-being, we find a number of more promising options. The literature here has largely followed Derek Parfit in classifying theories of well-being into three types: hedonist, objective list, and desire satisfactionist.[5] Given the prominence of this classification of theories, it will be worthwhile to briefly discuss each type of theory and its corresponding problems. This will pave the way for the arguments to come.

Consider first, then, *hedonistic* theories of welfare. These theories are characterized by their reduction of welfare to pleasurable or enjoyable experience. In its simplest form, *hedonism* claims that well-being consists in the experience of pleasure; that ill-being consists in the experience of pain; and that the more intense, or longer-lasting, an experience of pleasure (pain) is, the better (worse) off its subject is thereby made. Since this view claims that pleasurable experience is the only kind of thing that is good for us, and painful experience is the only kind of thing that is bad for us, it is monistic. It seems equally clear—at first glance, at least—that the theory is non-hybrid. Anything that is good for us, on this theory, is good for us precisely because it is an experience of pleasure. Whether hedonism counts as subjectivist or objectivist is, however, less clear. This depends (at least in part) upon the nature of pleasure.[6] On the one hand, we might take pleasure to consist in a

[5] Parfit 1984 (appendix I).

[6] The following also brings out a potential worry for our current understanding of hybrid theories. If the type(s) of thing that a theory claims can contribute to a subject's well-being can, on reflection, be reduced to some more basic concepts, the theory may in effect "become" a hybrid theory. The worry is then that it might turn out that near every theory of well-being on offer is a hybrid. And that might seem implausible. But, then again, it also might not.

distinct *phenomenology*, or positive feeling tone, which may be experienced regardless of any of our attitudes. In this case, hedonism would presumably count as objectivist: what is good for us (the experience of pleasure) will not owe its goodness to the fact that we have some attitude toward it. On the other hand, we might take pleasure to consist (at least in part) in a special kind of *attitude*, whose object just is the relevant feeling or sensation. In this case, so long as our theory claims it is the *experience* of the attitude (such as the experience of occurrently desiring some sensation), rather than the *object* of the attitude (such as the particular sensation that we occurrently desire), that has prudential value, our theory will again count as objectivist.[7] Indeed, it is only if our theory claims that this value resides in the object of the relevant attitude that it will count as subjectivist. Yet, given the current characterization of hedonism, this theory would not also count as hedonistic. This is because it would take the object of the relevant attitude, such as the sensation itself, to be what was basically good for us, while also claiming that pleasure was essentially (at least in part) that attitude, meaning that something besides (and more basic than) pleasure was basically good for us. It seems, therefore, that any version of hedonism that is worthy of its name will be objectivist.

While simple, hedonism captures a number of important intuitions about what makes a life go well. For one, it can make sense of the thought that a life without any enjoyable experiences at all could not be a good one. So long as well-being consists in enjoyable experience, a life without enjoyment is a life without welfare. For another, it has

[7] There are, then, actually two versions of the simple hedonist view. On the first, all and only experiences of *feeling* pleasure or enjoyment are non-instrumentally good for us. A theory of (roughly) this kind has been defended by Roger Crisp (2006, ch. 4). On the second, all and only experiences of pleasure-constituting *attitudes* are non-instrumentally good for us. A theory of (roughly) this kind has been defended by Fred Feldman (2004, ch. 4). I'll have more to say about this in chapters 7 and 8.

the plausible implication that the worst kind of life would be one characterized entirely by painful experiences: if ill-being consists in the experience of pain, it follows that nothing could be worse for us than a life full of torture. Furthermore, the theory offers a simple explanation as to why very many things that seem to be good for us are in fact good for us. Eating tasty food, spending time with close friends and loved ones, reading great books, and developing our physical and psychological capacities are all things that we tend to enjoy. They are also things that, in most cases, we take to be good for us. Hedonism vindicates this second thought, simply enough, by appealing to the first: the multitude of things we take to be good for us are good for us precisely because, and to the extent that, they are enjoyable or pleasurable.

Still, this way of explaining the goodness of these activities seems to miss out on something important: their goodness does not seem to derive entirely from their pleasurableness. This point is brought out perhaps most forcefully by Nozick's experience machine thought experiment.[8] Faced with the option of entering a machine that would simulate the experiences of a life replete with such good and pleasurable activities, or of actually having these same experiences in the real world, it seems clear that most of us would prefer the latter. If, in addition, we had to give up our life in the real world in order to experience the same life on the experience machine, our preference here would presumably only grow stronger. Since the experiences in either case would be identical, however, simple hedonism cannot make sense of this preference: the theory implies that the lives would be equally good. That's because the two types of life would be identical in terms of enjoyableness, and would therefore contain the same amount of non-instrumental goodness.

[8] Nozick 1974 (pp. 42-45).

Of course, the hedonist is free to claim that our intuition here is simply leading us astray, but to many of us it will seem overwhelmingly plausible that being massively deceived, or failing to experience pleasure in a way that properly links us to the world, can adversely affect our level of well-being. The experience of pleasure does not seem to be the be-all-end-all of the goodness of a life. Whether these experiences are had in the "right way," whatever this way might be, also seems to matter.[9] If that's right, the simple hedonist view must be false.

Consider next desire satisfactionist theories of welfare. These theories claim that a subject's well-being consists in the satisfaction of certain of the subject's desires. Again in its simplest form, *desire satisfactionism* says that well-being consists in the satisfaction of any of a subject's intrinsic desires.[10] So what is good for us, on this theory, is our getting what we intrinsically want, whatever that might be. Since this theory claims that the satisfaction of our desires is the only type of thing that is non-instrumentally good for us, it is plainly monistic. And since it restricts the desires that might contribute to our welfare—counting the satisfaction of only our intrinsic desires as constituting our well-being—it seems it is (technically) also hybrid. Finally, unlike with simple hedonism, the matter of whether simple desire satisfactionism is objectivist or subjectivist seems to admit of a clear answer: it is objectivist. This is because it says that a thing is a constituent of a subject's well-being if and only if and because that thing is the satisfaction of one of the subject's intrinsic desires. It therefore does not take the objects of the subject's attitudes alone to determine what is basically good for them.

[9] I will offer two suggestions of what the "right" way is in chapters 4 and 7.

[10] Admittedly, an even simpler version of desire satisfactionism is available: it claims that well-being consists in the satisfaction of just any of the subject's desires. This theory is clearly implausible, however, because it implies that my purely instrumental desire to (for instance) go to the dentist makes my extremely unpleasant time at the dentist basically (not just instrumentally) good for me, which does not seem particularly plausible.

Even this simple version of desire satisfactionism has a number of attractive features. Recall what pushed us away from hedonism: it could not account for the fact that whether our experiences of pleasure connect to the world in the "right way" might affect our level of well-being. Desire satisfactionism does not face this worry, since it claims that the proper connection between our desires and the world is precisely what well-being consists in. That is to say, that our desires connect to the world in the "right way"—that they are satisfied—is precisely what is good for us, according to desire satisfactionism. So the theory does not imply, for example, that a life in the real world would be just as valuable as a life with the same experiences in the experience machine, since very many of one's intrinsic desires in the experience machine would go unsatisfied in the real world. The theory thus ensures a proper link between our welfare-constituting attitudes and the world. Furthermore, since the theory is subjectivist, desire satisfactionism secures another important connection—one between what is non-instrumentally good for us and what we care about or are attracted to. It seems plausible that what is basically good for us must be something to which we are, or would under certain conditions be, genuinely attracted. After all, if it is claimed that something is basically good for us, and yet we do not find this thing attractive, it will always be possible for us to be indifferent toward the thing—or, worse yet, for us to *hate* it; in which case, the thing will not seem to be basically good for us after all. This claim about the connection between what is good for us and our motivations is sometimes called *internalism* about well-being.[11] Since desire satisfactionism claims that well-being consists in the satisfaction of one's desires, it follows that what is good for one here must be determined by what one

[11] It is also sometimes called the "resonance constraint." Here is Peter Railton's seminal expression of the view: "what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive … It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him" (1986, p. 9).

intrinsically desires. Accordingly, the theory also seems to be motivated by this plausible internalist view.

But of course desire satisfactionism, too, seems to face decisive objections. The first concerns the putative support desire satisfactionism receives from internalism about well-being. The problem is that we have many intrinsic desires whose objects do not genuinely appeal to us in the way internalism requires. It might be, for example, that a subject intrinsically desires to read some book, and yet they find reading unpleasant. They are not genuinely attracted to reading the book—and, when they go on to read it, they unsurprisingly dislike every second of doing so. A life full of satisfied intrinsic desires of this sort would not seem to be a good one. Yet the simple desire satisfactionism under consideration would judge such a life to be going quite well: one would constantly be getting what one intrinsically wanted. Since one would not in any sense be genuinely attracted to the things constituting one's welfare, desire satisfactionism runs afoul of, not just the intuition that this life would not be a good one, but also internalism about well-being. In addition, this example brings out a second problem for desire satisfactionism: we have many intrinsic desires whose satisfaction intuitively makes no difference to how well our lives go. We intrinsically desire things that we dislike, that we rarely (if ever) think about, that we do not know obtain, that are based on false beliefs, that are irrational, that concern other people's well-being, and so on. When we have these desires, and we are not genuinely attracted to their objects, their satisfaction will intuitively make no difference to our level of well-being. Simple desire satisfactionism cannot account for this intuition, since the theory claims that well-being consists in the satisfaction of any of our intrinsic desires. A third and final objection to this theory concerns one of these types of desire: those the subject does not know obtain. These desires present a special problem for desire satisfactionism, because they

reveal that what matters to well-being is not, specifically, the *satisfaction* of a subject's intrinsic desires. After all, if subjects seem better off, not when their desires are satisfied, but instead only when they are aware of the fact or believe that their desires are satisfied, it follows that well-being cannot consist entirely in the satisfaction of their desires. If, for example, Parfit's desire that the stranger he meets on a train has a good life is satisfied, and he does not know it, he does not seem to be any better off.[12] This suggests that only desires that one believes obtain can make one better off. And this suggests that the *satisfaction* of desires, in itself, is not the source of non-instrumental goodness. The current version of desire satisfactionism should therefore be rejected.

Consider finally objective list theories of well-being. Theories of this sort propose a list of basic objective goods—or goods that are said to be, in themselves, worth having or pursuing—and claim that well-being consists in the possession of those goods.[13] Proposed lists often include things like knowledge, autonomy, pleasure, and friendship. While near all objective list accounts that have been legitimately proposed and defended have included more than one item on their lists—and have therefore been pluralistic—the possibility of a monistic objective list account cannot be rejected out of hand. So long as a theory claims that a thing is a constituent of a subject's well-being if and only if and because that thing is an objective good, the theory is an objective list account. If the theory claims that there is only one such good (such as knowledge), its status here is unaffected. Nonetheless, most all objective list accounts propose more than one objective good, and are thus pluralistic.

---

[12] Parfit 1984 (p. 468).

[13] Here it might be wondered what is meant by talk of a good's *possession*. As I see it, it is likely just a convenient of way of expressing that the subject shares the proper sort of connection to the good. So, for instance, autonomy is possessed when the subject is autonomous; knowledge is possessed when the subject knows. If, at any rate, there is a deeper issue for the objective list approach arising here, it will not be raised by me.

Whether objective list theories count as hybrid or non-hybrid depends, of course, on the nature of the good or goods they propose. If at least one such good is such that the contribution it makes to a subject's well-being in part depends on whether, and to what extent, the thing has some further property, the theory as a whole will count as hybrid. Otherwise, the theory will be non-hybrid. In contrast, whether an objective list theory counts as objectivist or subjectivist does not depend on any features of the proposed good or goods. Since objective list accounts claim that a thing is a constituent of a subject's well-being if and only if and because that thing is an objective good, it follows that any such account will identify the constituents of a subject's well-being by reference to an independent evaluative standard. Consequently, objective list accounts are invariably objectivist.

In fact, their objectivism is plausibly their greatest selling point. As we have seen, one problem with desire satisfactionism is that we have many desires whose satisfaction does not intuitively seem to make us any better off. This seems to show us that, from the perspective of our well-being, we can desire the wrong things. Yet this seems to suggest that we can also desire the *right* things—or that there are some other things that it is right for us to desire. And these are, presumably, just the things that would be objectively good for us to have. Thus what leads us away from simple desire satisfactionism seems to lead us toward objectivism, and more specifically the objective list approach. Similarly, our issue with hedonism was its implication that a life experienced on the experience machine would be just as valuable as the same life experienced in the real world. Depending on which goods a theory proposes, and how these goods relate to each other in constituting one's welfare, an objective list theory may easily evade this issue. It might claim that there are things other than pleasure that help determine how well a life goes, such as knowledge or friendship, and

that a life on the experience machine would not rate well along these other dimensions. Hence objective list accounts can pass a number of tests that their competitors fail. More fundamentally, and as these examples bring out, objective list theories can account for the plausible thought that more than one kind of thing can be non-instrumentally good for us. (Of course, so can any pluralist theory, but objective list accounts are often characterized as being motivated specifically by this thought.) Pre-theoretically, it might seem clear that what is good for any subject is not restricted to the mere experience of pleasure or satisfaction of desire, but instead comprises a much broader range of things, including being healthy, developing certain physical and psychological capacities, accomplishing certain goals, and doing something worthwhile in life. Rather than reducing the goodness of each of these things to the pleasure or desire-satisfaction that they may involve, objective list theories aim to give a richer account by recognizing each of these things as non-instrumentally valuable in their own right.

Since there are very many possible objective list accounts, it is of course difficult to offer any single objection to them all. Nonetheless, one objection to many of those that have been proposed and defended is that they run afoul of the internalist intuition that something cannot contribute to our well-being unless we are in some sense attracted to or pleased by it. If such a theory does not include on its list pro-attitudes (such as desires or pleasures) of any sort, it will imply that a life containing the proposed objective goods yet experiencing nothing but pain and misery would nevertheless be a good one. And even if a theory includes pleasure or the satisfaction of desires on its list, it will likely imply that the other things on the list—since they are objective goods—can contribute to one's well-being even in the absence of any desire for or pleasure taken in them. Again, however, a life with no pleasure or attraction—or, worse yet, one full of misery—could not be made good by

friendship and autonomy alone. Objective list accounts may thus face charges of failing to properly connect the things that are non-instrumentally good for us to our motivations. Of course, it might instead be claimed that certain pro-attitudes are *constitutive of* the proposed objective goods, which means that anything constituting our well-being will be something we are properly attracted to.[14] Yet, in that case, it will likely seem more plausible that it is the pro-attitudes *themselves*, rather than the putative objective goods of which they are partially constitutive, that possess non-instrumental value. It seems, then, that the objective list theorist can meet the demands of internalism only by giving us reason to doubt the plausibility of their theory.

Our discussion to this point might seem somewhat discouraging. Having seen many of the main points for and against the three classic types of well-being theory, we are left at an impasse. At the very least, few would say that the matter of which type of theory should be preferred admits of an obvious answer. But I do not think that we should be discouraged. In fact, I think that there is an important lesson to be learned from our discussion. In particular, as I see it, each of these theories seems to contain at least a kernel of truth—or something about the nature of well-being that they are getting right. While each theory faces decisive objections, it seems to me that each captures an important datum about the nature of well-being. It is no accident that Parfit chose just these three types of theory, and that people have largely followed him in classifying theories in this way. No doubt this tripartite division of theories is in one sense arbitrary (it is not exhaustive, for one). But what justifies our continued use of it is that it ensures that our theories will approximate the truth. Each type of theory has got something going for it. Yet each faces its own fundamental problems.

[14] A view of this sort has been proposed by Guy Fletcher (2013).

All of this points to a promising way forward. If each of these types of theory are getting something right, but cannot by themselves get everything right, then we might try to amalgamate them in a way that gets us both the strengths of each and the weaknesses of neither. Perhaps debating the truth of these three types of theory is not unlike debating which piece of the puzzle depicts the puzzle in its entirety. Of course none of them alone capture the true nature of well-being—that's just what we should have expected. Still, each tells us something important about the nature of well-being. And perhaps together they can tell us everything.

The idea, then, is that we might take more than one of the three classic types of well-being theory—hedonist, desire satisfactionist, and objective list—and combine them to create a hybrid theory. This approach seems promising, because it offers the prospect of pitting the strengths of each type of theory against the weaknesses of the other. The hope is that this would leave us with a theory of well-being that gives us everything we want from any such theory.

In recent years, a number of philosophers have adopted this approach, and constructed hybrid theories of well-being. As it happens, many of them have focused on two of the three classic types of well-being theory in constructing these theories. Specifically, they have focused on integrating features from hedonist and objective list theories of well-being. It seems clear, in light of our discussion of each of these types of theory, why these types in particular should seem promising candidates for a plausible hybrid theory. Many of the problems with hedonist theories seem apt to be addressed by objective list theories, whose own problems seem apt to be addressed by hedonist theories. The hedonist's greatest obstacle is their inability to account for the intuition that whether our experiences of pleasure connect to the world in the "right way" can affect our level of well-being. The

objective list theorist's greatest obstacle is their inability to account for the intuition that whether our motivations connect to our possession of (putative) objective goods can affect our level of well-being. It seems plausible, then, that these obstacles might be avoided by making well-being a matter of, not just experiencing pleasure or possessing objective goods, but experiencing pleasure taken in objective goods. This would presumably ensure that our experiences of pleasure connected to the world in the "right way," and that the objective goods constitutive of our well-being were things to which we were properly attracted.

That, at least, is the hope. A number of people, as I say, have recently converged on this thesis. Though the specifics of their accounts vary—and, in many cases, go missing entirely—the general idea is the same: well-being consists in *enjoying the good*. In the following chapter, I'll review some of these accounts, and offer a closer look at their underlying structure. While there are a number of ways of formulating views of this type, I'll argue that most all of them lead to implausibility. Still, there are at least two ways of formulating them that, I'll suggest, are immune to this charge. Finally, I'll argue that one of these two ways is to be preferred to the other. This will leave us with a framework for constructing a plausible hybrid theory.

2 Enjoying the Good

One way of reaching the plausible claim that well-being consists in enjoying the good goes roughly as follows. A life without pleasure cannot be a good one; thus, pleasure is necessary for well-being. Yet experiencing a great deal of pleasure (and little pain) does not guarantee that a person's life goes very well—the person may (e.g.) have spent their life on the experience machine. So, pleasure alone is not sufficient for the best sort of life. The possession of objective goods is similarly insufficient—such possession must in each case be accompanied by something like pleasure. But perhaps possession of these goods is, again like the experience of pleasure, also necessary for well-being. After all, pleasures taken in things that are not objectively valuable, or that are objectively disvaluable, often seem not to make us any better off. This is true of both "false" pleasures, which would pervade one's life on the experience machine, and "base" pleasures, such as the pleasure one might take in watching an innocent person being tortured. If we wish to deny that these experiences of pleasure can make a subject any better off, we might claim that what is non-instrumentally good for us is, not experiencing just any pleasure or possessing just any objective goods, but taking pleasure in the objective goods we possess. On this approach, both experiencing pleasure and possessing objective goods are necessary components of well-being, yet they are only sufficient for well-being when taken together. Somewhat more precisely, the thought here is

Enjoying the Good: A subject's well-being can increase if and only if (i) the subject experiences pleasure and (ii) that pleasure is taken in an objective good.

As I've mentioned, there has recently been a kind of convergence on this view in the literature. It, or something quite like it, has been defended by Robert Adams, Neera Badhwar, Stephen Darwall, and Shelly Kagan, among others.[15] While the precise details of their accounts vary, Enjoying the Good serves as a general expression of their shared thesis.

Notice, though, that this formulation of the thesis is importantly deficient: it is neither enumerative nor explanatory. That is, it does not tell us *which* things constitute a subject's welfare, or *why* those things in particular so constitute it. Of course, this may not seem all that difficult a problem to solve. Take, for example, Christopher Woodard's formulation of the view, which we can call

EG0: For any subject S and any thing X, X is a constituent of S's well-being if and only if and because (i) X is an objective good and (ii) S takes pleasure in X.[16]

This definition tells us both which things constitute a subject's welfare—whatever is an X—and why those things in particular so constitute it—because they satisfy (i) and (ii). So, problem solved, right?

Well, not quite. Though this definition appears to solve our first problem—it is both enumerative and explanatory—it does so only by giving rise to a second. In particular, the

[15] Adams 1999, Badhwar 2014, Darwall 2002, Kagan 2009, Kraut 2007. Kraut abandons this thinking, however, in Kraut 2018.

[16] This is slightly adapted from Woodard 2015 (p. 7).

definition fails to tell us in what, specifically, well-being consists. It gives us some idea of what makes something good for someone, of course, but it leaves open which of the two conditions is more fundamental to this explanation. Indeed, it merely offers two necessary conditions that a thing must meet to basically constitute our well-being, and offers no explanation as to how these conditions ultimately relate, or how our well-being is ultimately constituted.

To see this, consider a more precise version of Enjoying the Good:

> *EG1*: For any subject S and any thing X, X is a constituent of S's well-being if and only if and because (i) X is an objective good, and only if (ii) S takes pleasure in X.

Notice that, unlike with EG0, the first condition here identifies the kinds of thing in which well-being consists—objective goods—whereas the second offers a necessary condition on those potential constituents of well-being. So while it is, on this approach, partly because a subject takes pleasure in an objective good that the objective good contributes to their well-being, it is ultimately the objective good itself that makes that contribution. The pleasure is not, strictly speaking, what constitutes one's well-being—it is not the source of prudential value.

That, at least, is what EG1 implies. But of course we might disagree: we might think that the experience of pleasure is itself the constituent of one's welfare, though its contribution is conditional on its being taken in an objective good; or that both the experience of pleasure and the objective good are the constituents, though the contribution

of each is conditional on the presence of the other.[17] Woodard's formulation of Enjoying the Good, while perfectly neutral, obscures all three of these more precise possibilities. And this is not a mere quibble. After all, Woodard's formulation leaves us with a theory that we cannot even classify in terms of our taxonomies from the previous chapter, since it is insufficiently precise. And whereas EG1 leaves us with a pluralistic, objective list theory of well-being, one of the more precise alternatives would leave us with a monistic, hedonistic one. As we have seen, this can have great implications in terms of the problems our theories face.

In any case, my point here is not that EG1 is more or less plausible than either of the two alternatives. It is instead that Woodard's discussion and formulation neglects the three more precise versions of Enjoying the Good, and that these theories are ultimately what we should focus on in assessing Enjoying the Good's plausibility. As it happens, I think that EG1 has just as much plausibility as each of the two more precise alternatives. But, unfortunately, I think that this amount of plausibility is none. Let me explain.

2.1 Three (failed) formulations

On Enjoying the Good, an experience of pleasure and the possession of an objective good are each individually necessary for an increase in well-being. But there are at least three ways that these conditions might be jointly sufficient. First, as EG1 suggests, it might be that the constituents of one's well-being are the objective goods one possesses, though these goods can only contribute to one's well-being so long as one takes pleasure in them. On this approach, pleasure does not, in itself, contribute to one's well-being. Rather, well-being

[17] All three of these alternatives are also mentioned by Kagan (2009, p. 261).

consists in the possession of objective goods—these goods are the sole constituents of one's well-being—though the goods will only make one better off provided one also takes pleasure in them. This is the view we are calling EG1.

EG1 faces a significant worry. In particular, it is difficult to plausibly explain why the relevant objective goods cannot contribute to a subject's well-being without the subject taking pleasure in them. After all, the view claims that prudential value resides in one's possession of the relevant objective goods. Yet, in that case, it is unclear why it should be that this value can only be realized so long as one also experiences pleasure, given such experience is, according to the theory, not itself prudentially valuable. That is, it is unclear why something that is not prudentially valuable—pleasure—should be necessary for the prudential value of some other things—objective goods—to be realized. Presumably, the objective goods' intrinsic prudential value-making features will be present regardless of whether a subject takes pleasure in them.[18] So why would this value only be "unlocked," as it were, by a subject's corresponding experience of pleasure?[19] If, for example, the theory includes knowledge on its list of objective goods, it will imply that while knowledge is a source of prudential value, that value can only be realized so long as the knowledge has pleasure taken in it. Yet the intrinsic properties of the knowledge will be the same regardless

---

[18] See Crisp 2006 (p. 123).

[19] Might proponents of EG1 say that, since pleasure is also an objective good, there is nothing implausible about the thought that it is needed for the prudential value of any objective good to be realized? Yes, but they would then seem to face the same problem in a different form. After all, if the prudential value of any objective good is realized only if one takes pleasure in the good, and pleasure is itself an objective good, it follows that the prudential value of pleasure can be realized only if it has pleasure taken in it. The prudential value of pleasure is then, on this approach, a fundamentally latent property—one that is "unlocked," oddly enough, by pleasure itself. Furthermore, the implication that experiences of pleasure make our lives go better only when we take pleasure in them seems obviously wrong. We can easily imagine a life where one would experience a ton of pleasure without ever taking pleasure in the fact that one had had these experiences. The current defense of EG1 implies that, if a life of this sort did not include any of the other objective goods, it would have no well-being—it would be no better than non-existence. And that seems implausible.

of whether a subject adopts this stance toward the knowledge. This approach then seems to make prudential value a somewhat mysterious, latent property of knowledge—one that lies dormant, coming about only if the knowledge has pleasure taken in it. But why should the presence of pleasure have anything to do with the prudential value of some other kind of thing? Given EG1's claim about what constitutes a subject's well-being, it does not seem apt to answer this question. This leaves us without an explanation as to why it would be, as this approach suggests, that prudential value was a latent property of objective goods.

Admittedly, the mere fact that it is unclear why something might be the case does not entail that it is not the case. We might lack a plausible explanation as to why pleasure is needed for the prudential value of objective goods to be realized, but still it might be the case that this accurately describes the nature of prudential value. But without any reason to think that this account—rather than some alternative—does accurately describe the nature of prudential value, it seems a strike against it that its truth would lack a plausible explanation. The idea that prudential value is a latent property of objective goods, waiting to be unlocked by a subject's corresponding experience of pleasure, does not seem particularly plausible. Thus if we can avoid offering an account on which prudential value is in this way a fundamentally latent property, and other things are equal, it seems we would be wise to do so.

On an alternative, sufficiently specific formulation of Enjoying the Good—our second—the constituents of one's well-being are experiences of pleasure, though whether these experiences contribute toward one's well-being depends on whether they are taken in objective goods. The objective goods then, on this approach, do not contribute to one's well-being. Instead, well-being consists in the experience of pleasure—such experience is the

source of prudential value—but experience of this sort will only make one better off provided the pleasure is taken in objective goods. More precisely, on what we can call

> *EG2*: For any subject S and any thing X, X is a constituent of S's well-being if and only if and because (i) X is a pleasure, and only if (ii) S takes X in an objective good.

Since it tells us that the only constituents of a subject's welfare are experiences of pleasure, EG2 is a form of hedonism. And since it tells us that there is just one kind of thing that can constitute a subject's welfare, it is monistic. In these respects, EG2 is quite unlike EG1.

There is another respect in which EG2 is unlike EG1. Specifically, EG2 does not imply that prudential value is a kind of latent property of the things that possess it. That is, the claim here that prudential value resides in pleasure does not lead us to the conclusion that this value is a potential property of any pleasure. This is because we can instead precisify this claim to state that prudential value resides in all and only pleasures of a certain kind—in particular, pleasures taken in objective goods. This was not an option on EG1, because EG1 claimed that prudential value resides in objective goods. Since any objective good is such that a person could take pleasure in it, our attempt to precisify the claim in that case left us to conclude that prudential value is a potential property of any objective good—a property that is "unlocked" by a subject's corresponding experience of pleasure. To put the point here slightly differently: whereas the phrase "objective goods that one takes pleasure in" does not pick out a proper subset of all objective goods—since one could, presumably, take pleasure in any given objective good—the phrase "pleasures taken in objective goods" does pick out a proper subset of all pleasures. As a result, whereas the first approach identifies prudential value as a potential property of any objective good, the current approach identifies it as a

property of all and only pleasures of a certain kind. EG2, accordingly, does not have any latency problem.

But it does have a problem of its own. In particular, EG2 simply cannot be made to fit with our intuitions about what makes for a good life. On this approach, only pleasures of a certain kind can make a subject's life go better. Thus any pleasure that is not of this kind cannot affect how well a subject's life goes. But, in that case, EG2 implies that a life containing just some amount of pleasure of this latter sort—pleasure not taken in an objective good—would be no better for a subject than never having lived at all.[20] If, for example, EG2 implies that the "false" pleasures felt in the experience machine would not contribute to one's level of well-being, it follows that a life containing only such pleasures would be no better for a subject than no life at all. To many of us this implication will seem extreme. Surely we would prefer living a life of false pleasure to just calling it quits (at least, other things being equal). EG2 cannot make sense of our preference here. Since each alternative would contain the same amount of pleasure taken in objective goods, each would contain the same amount of prudential value. So, either way, we would be equally well off. This, I think, is implausible. It seems to me more plausible that pleasure taken in things that are not objective goods may contribute at least a small amount to a subject's welfare. That's because a life containing only pleasure of this kind seems better than no life at all, especially if the pleasure would be very great in quantity. Thus if I am correct in my thinking here, it cannot be the case that experiences of pleasure contribute to a subject's well-being just so long as they are taken in the objective goods the subject possesses.

---

[20] Admittedly, it might seem odd to compare the welfare level of a welfare subject with the welfare level of someone who is not a welfare subject, and who therefore cannot have a level of welfare. Yet this is not what I take myself to be doing. Rather, I'm comparing the welfare level of a welfare subject with the welfare level of a welfare subject (who therefore can have a level of welfare) if they did not exist.

There is a third way of formulating Enjoying the Good. This approach dispenses with the idea that one of the two conditions in Enjoying the Good has a type of primacy over the other. That is, it rejects the thought that one of these conditions specifies the kind of thing in which well-being consists, while the other merely specifies the conditions under which things of that kind can constitute one's well-being. On this approach, both the possession of objective goods and the experience of pleasure contribute to one's well-being, though whether either of these will make such a contribution depends upon the presence of the other. More precisely, on what we can call

> *EG3*: For any subject S and any thing X, X is a constituent of S's well-being if and only if and because either (a) (i) X is an objective good, and only if (ii) S takes pleasure in X, or (b) (i) X is a pleasure, and only if (ii) S takes X in an objective good.

This formulation of Enjoying the Good is, in effect, a combination of our first two formulations. It tells us that well-being consists in both taking pleasure in the good, and possessing objective goods that one takes pleasure in.

Since it is a combination of EG1 and EG2, however, EG3's fate has already been decided: it is unacceptable. This is because EG3 adopts each of the problems that led us to reject EG1 and EG2. Since it takes on the claims of EG1, EG3 implies that prudential value is a latent property of objective goods; and since it takes on the claims of EG2, it implies that taking pleasure in things with little or no objective worth is no better for us than ceasing to exist. EG3 is therefore even less plausible than either of the formulations it attempts to combine.

2.2 Two additional formulations

The problem with EG2 (and, thus, EG3) is that it picks a set of pleasures, and claims that these pleasures alone have prudential value. But, intuitively, it seems that other pleasures also have this value—that they can also make a life go better. In fact, it seems that *every* pleasure has at least *some* of this value: any experience of pleasure would be better than non-existence. By restricting the set of pleasures that have prudential value to just those taken in objective goods, EG2 neglects this important point.

Recall that we began our discussion of Enjoying the Good with the plausible suggestion that both pleasure and objective goods are necessary for an increase in well-being. What motivated the latter part of this suggestion—that objective goods are necessary—was the thought that pleasure cannot be the be-all-end-all of prudential value. The pleasures experienced while on the experience machine, for example, may seem less prudentially valuable than those same pleasures experienced in the real world. So, the thought went, perhaps these pleasures must meet some independent standard in order to make us better off. As we have seen, this approach fails. Notice, however, that the divergence in value between pleasures of the same intensity and duration experienced on and off the experience machine does not imply that pleasure is not the be-all-end-all of prudential value. Rather, it implies just that the prudential value of an experience of pleasure cannot be determined by the intensity and duration of the experience alone. It is this portion of simple hedonism that the experience machine seems to disprove. Perhaps if we thought that lives lived on the experience machine could not have any well-being, we could conclude that there must be something more to well-being than experiences of pleasure. After all, if we thought this, we could note that while some lives are good for those experiencing them in the real world, the

same lives would have no value if they were experienced on the experience machine, and thus well-being cannot consist in the experience of pleasure. Far from this, however, what we have learned from rejecting EG2 is that, in fact, *every* pleasure has at least *some* non-instrumental value. The experience machine thus teaches us, not that only some experiences of pleasure are good for us, but that *how* good *any* experience of pleasure is for us cannot simply be a function of its intensity and duration. Our lesson is that there is more to the value of pleasure than its intensity and duration; not that there is more to well-being than pleasure.

It may at this point be objected: it is simply not clear that all experiences of pleasure are good for us. We have considered pleasures taken in things with no objective worth, but not pleasures taken in objectively disvaluable things. Consider, for instance, pleasure derived from torturing babies. While some may have the intuition that experiencing pleasure of this sort is at least somewhat good for the subjects who experience it, others may not. They may find such pleasure simply detestable, and devoid of any goodness.

Is there anything that can be said in response to these people? I think there is. For any pleasure, including that derived from torturing babies, we can imagine, on the one hand, a life in which one experiences this pleasure at every moment from beginning to end, and, on the other hand, non-existence. We can even make this life infinitely long, and this pleasure maximally intense. We can then ask: which seems better, the life of constant pleasure, or non-existence? As I see it, there is no pleasure, not even that derived from torturing babies, for which our answer to this question could be Non-existence. Regardless of how detestable the source of a pleasure seems, experiencing that pleasure continuously, for eternity, will always seem better for the subject of the experience than non-existence. To claim otherwise is, I think, more implausible than to claim that all pleasurable experience is

at least somewhat basically good for us. Yet, if that's right, and every pleasure is such that experiencing it continuously for eternity is better for the subject who experiences it than experiencing nothing, it at least follows that every pleasure is such that there are conditions under which it is at least somewhat basically good for us. And the most plausible explanation for this, it seems clear to me, is that every pleasure is at least somewhat basically good for us. After all, it is unclear how this basic goodness could arise when the pleasure is maximally intense and long-lasting, unless it were also present to some extent whenever the pleasure was experienced. I do not see a more plausible explanation available here.[21] It seems, then, that every pleasure is at least somewhat basically good for us.

If all that is right, and there is more to the value of a pleasure than its intensity and duration, then what feature of pleasure does simple hedonism neglect in assessing a pleasure's value? Our earlier attempt to modify the view suggests that the answer here is the pleasure's *object*. The modified view, EG2, claimed that only pleasures whose objects are of a certain kind—those with objective worth—can contribute to a subject's well-being. But this view misdiagnosed the problem with simple hedonism. It is not that simple hedonism implies that too many pleasures possess prudential value. It is that simple hedonism fails to recognize the third determinant of a pleasure's value. The view cannot account for the plausible thought that two subjects whose experiences of pleasure are identical with respect to their intensity and duration, but distinct in terms of their objects, may not have identical improvements in welfare. The experience machine offers one reason for thinking this, but examples can be multiplied. If, for example, one subject takes pleasure in the fact that an

[21] Of course, it might be suggested that there is a threshold at which an experience of pleasure becomes sufficiently intense or long-lasting to count as basically good, but any such threshold would surely seem arbitrary.

innocent person is being tortured, while the other takes pleasure of the same intensity and duration in the fact that the world is at peace, the other would seem to experience a greater increase in welfare than the one. Notice again that this does not imply that objective goods are *necessary* for an increase in well-being: we have not said that certain pleasures would contribute *nothing* to one's well-being. Rather, we have said that some pleasures would make a greater contribution than others, even if their intensities and durations were the same. This suggests that the contribution an experience of pleasure makes to one's welfare depends, not only on its intensity and duration, but also on its object.

It seems, then, that there is an alternative way forward. We might claim, not that objective goods are necessary for an increase in well-being, but that the extent to which an experience of pleasure contributes to one's well-being is determined, in part, by the extent to which the object of one's pleasure is objectively valuable. On this approach, objective goodness *enhances* a pleasure's contribution to one's welfare.[22] It does not act as a constraint on a pleasure's ability to make such a contribution. Since this view adds an objective component to simple hedonism, we can call it *objective hedonism*. Like simple hedonism, it claims

[22] Here it might be wondered to what extent objective goodness enhances a pleasure's contribution to one's welfare. Might pleasure taken in objective goods be *incommensurably* better than pleasure taken in other things—that is, might there be some amount of pleasure of the first sort (where this amount might be just any) that would contribute more to one's welfare than any amount of pleasure of the second sort? I am inclined to think not. It seems to me excessive to limit the value of pleasures taken in things that aren't objective goods in this way. I also see little reason to think that such a threshold in value exists.

For any subject S and any thing X, X is a constituent of S's well-being if and only if and because X is an experience of pleasure had by S.

But it diverges from simple hedonism by adding

The extent to which X contributes to S's well-being depends on the duration and intensity of X, as well as the objective worth of the object of the pleasure in X.

Since it claims that all and only experiences of pleasure constitute a subject's well-being, objective hedonism is indeed a form of hedonism. And since it claims that there is only one kind of thing that can constitute a subject's well-being, objective hedonism is monistic. Given that there are a number of ways of making this view more precise—depending on what we take each of pleasure and objective worth to consist in—objective hedonism may be seen as a more general approach, or as a type of theory, that we may agree upon even if our precise formulations of it are at odds.

Likewise, if we prefer the objective list approach, we might claim that the extent to which an objective good contributes to one's welfare is partly determined by the extent to which one takes pleasure in that objective good. Pleasure, on this alternative view, enhances an objective good's contribution to one's welfare. Again, it does not act as a constraint on an objective good's ability to make such a contribution. So this view claims

For any subject S and any thing X, X is a constituent of S's well-being if and only if and because X is an objective good had by S.

And

The extent to which X contributes to S's well-being depends (at least in part) on both the objective worth of X, and the extent to which S takes pleasure in X.

This is, again, perhaps more properly conceived as a type of theory, since any objective list account could presumably adopt this structure.[23] Whether theories of this type will have certain other features—e.g., whether they are monistic or pluralistic—will then depend more on the contents of the particular theories than on the fact that they adopt this structure.

Importantly, notice that each of these approaches are silent on whether the proposed constituents will make any contribution to one's well-being when they lack the additional value-determining property. Each says that the extent to which a constituent will contribute to one's welfare partly depends on the extent to which the constituent instantiates some further property, but this tells us nothing about the contribution the constituent will make if it lacks that property. This is important because it means that these views may avoid the challenges faced by EG1-EG3 above. More specifically, unlike EG1-EG3, these views can claim that the proposed constituents will have at least some prudential value when they lack the relevant further property. And so, again unlike EG1-EG3, these views can account for the fact that a pleasure whose object is not objectively good nevertheless seems to contribute

[23] Richard Kraut's (2018) "experientialism" appears to be one theory of this sort.

at least some prudential value to a life; and that an analogous claim may seem true of objective goods. It seems, therefore, that we will want to formulate these views in a way that allows them to account for these intuitive judgments: we will want to say that the proposed constituents will have at least some prudential value when they lack the relevant further property.

We are left, then, with two potential views. On the first, all and only experiences of pleasure have prudential value, and the amount of prudential value had by a given experience of pleasure is determined by the pleasure's duration, intensity, and object—the longer or more intense the experience of pleasure, or more objectively valuable the pleasure's object, the more prudential value the experience has. On the second, all and only objective goods (whatever those might be) have prudential value, and the amount of prudential value had by a given objective good is determined, at least in part, by the amount of pleasure that is taken in that good. (To keep things simple, we will remain neutral on the other determinants of an objective good's prudential value.)

Now, it may be thought that this latter view suffers from just the same issue as the earlier view that identified the constituents of well-being as objective goods that a subject takes pleasure in—or, EG1. More specifically, it may be objected that there is no plausible explanation as to why something that is not prudentially valuable—pleasure—might affect the amount of prudential value that some other things—objective goods—have. After all, the objective goods' prudential value-making features will presumably be present to the same extent regardless of whether one takes pleasure in them. It then seems that we lack any plausible explanation as to why pleasure would play the role that this theory assigns it.

In this case, however, the objection can be easily avoided. Notice that the objection hinges on the claim that, on the current view, pleasure is not prudentially valuable. But of

course we have not yet identified which things are objective goods, and thus which things have prudential value. To avoid the current objection, then, all we must do is claim that one of these things is pleasure. So long as pleasure is itself an objective good—and so a source of prudential value—its role as a partial determinant of any objective good's prudential value will seem fully justified.[24]

But this is, of course, no reason to prefer this view to the one we have called objective hedonism. Have we any reason to prefer one of these views to the other? Well, notice that the objective list approach, as we have now developed it, must be pluralistic. This is because we've already said that this list must include pleasure, and failing to expand the list from there would effectively leave us back with simple hedonism. So, this view tells us that there is more than one kind of thing that makes something prudentially valuable. Objective hedonism, on the other hand, tells us that there is just one prudential value-making property—the property of being a pleasure. Thus if we have no other reason to prefer one of these theories to the other—and it seems to me that, as things stand, we don't—we might base our preference on the fact that one is pluralistic and the other is monistic. A reason to prefer one of these types of theory to the other would serve as a reason to prefer a token of the one to a token of the other. Have we any reason of this sort?

We do, I think. This is simply that monistic theories are more parsimonious than pluralistic theories. It is, of course, often seen as an advantage for views of the latter sort that they can account for the (supposed) prudential value of a greater range of things. They can account, for example, for the putative prudential value of pleasure, knowledge, autonomy, and so on, whereas monistic theories must stick to attributing this value to just one type of

[24] To see why this move is not also available to proponents of EG1, see footnote 19.

thing. But this advantage becomes a weakness when monistic theories can account for just as much and with fewer resources. This is the case, I take it, with the two approaches we're currently considering. So long as the objective worth of the objects of one's pleasures correspond to the objective goods that one might take pleasure in, these theories will likely yield identical results in most any case. If that is right, it seems objective hedonism should be preferred.

To be sure, this is not the strongest argument one could give for preferring one theory to another. But when the theories are as similar as the ones we are currently considering, this may be the best we can hope to do. Even so, let me try to offer one more point in favor of objective hedonism. There is one kind of case, in particular, in which the two theories as formulated must reach distinct results. Since the objective list approach counts pleasure among its goods, both theories will imply that even a pleasure whose object is not objectively valuable can make at least some contribution to a subject's well-being. On the other hand, only the objective list approach will imply that even objective goods that a subject does not take pleasure in can make at least some contribution to the subject's well-being. (This amounts to the rejection of the claim that pleasure is necessary for an objective good to contribute to one's well-being—or, EG1.) Thus, to separate the two views, all we must do is imagine a subject who possesses some (putative) objective good, but who never receives any pleasure from it. Now, since we have not specified any of the goods on this theory's list besides pleasure, we do not yet have the resources to imagine this. But, of course, we can speculate as to what these other goods might be. As we have seen, goods commonly appearing on lists of this sort include knowledge, autonomy, and friendship. For at least some of these goods, it may be difficult to imagine a case where a subject who possesses them never experiences any pleasure in connection with them. Friendship might

be like this. If our subject never has a single enjoyable experience in the course of developing and maintaining a putative friendship, we would likely be tempted to react either by doubting that their relationship was actually one of friendship, or else by implicitly sneaking in some experiences of pleasure in our conception of the case. Still, some of the listed goods are not difficult to imagine as being possessed without any corresponding experience of pleasure. Knowledge might be like this. It seems we can easily imagine a case where our subject has some piece of knowledge—say, that 6 + 10 = 16—but they never experience any pleasure in connection with this knowledge. They neither like nor dislike having this piece of knowledge; they're quite indifferent about the whole thing. They almost never think about it, but when they do, the fact that they have this piece of knowledge does nothing to move them—they just don't care. The subject's attitude is then not unlike the attitudes we ourselves may adopt toward trivial facts that we're aware of, such as the fact that the car that I've just seen looked white, or that I have *x* number of hairs on my head. This attitude is, of course, pure indifference.

Is this person made any better off by having this piece of knowledge? It seems to me that they are not. The knowledge may no doubt have some instrumental value, but if it never does anything to genuinely attract or please our subject, it does not seem to me that it could, in itself, make our subject any better off. If our subject had many pieces of knowledge of this sort, and never had a single experience of pleasure throughout their life—in connection with the knowledge or otherwise—again it seems to me that this knowledge would have done nothing to make their life go better. While the objective list approach need not imply that the subject's life in this case would have been a very good one, it does imply that a life of this sort must have some non-instrumental goodness. More precisely, it implies that there is some objective good that can contribute to a subject's well-being even if the subject never

has a single experience of pleasure. (Notice that it does not imply that some good can contribute to one's well-being even if one is made *miserable* by it, since, presumably, the view also claims that the extent to which a good contributes to one's *ill*-being is in part determined by the extent to which one is made miserable by the thing.) Though the defender of this approach may object that knowledge does not appear on their list of goods, the problem remains for any good that appears on their list besides pleasure. It may be difficult to imagine a case where the relevant good is had without any corresponding experiences of pleasure, but, as I see it, careful analysis of any such case will reveal that the putative good alone does nothing to improve a subject's well-being. If that is right, it follows that this objective list approach is false.

There is a sense in which my argument against the objective list approach here is an appeal to the internalist intuition that a thing cannot contribute to a subject's well-being unless the subject is in some way attracted to or pleased by it. It seems that objective list theories cannot account for this intuition without claiming that attraction or pleasure is a necessary condition for an objective good to contribute to one's welfare. After all, as we have seen, if they do not claim that attraction or pleasure is necessary, it will be possible to have an objective good whose possession intuitively makes one no better off, but according to the theory contributes to one's welfare. This is the case, even if the theory claims that pleasure or attraction enhances the contribution an objective good makes. As soon as the objective list theorist concedes that attraction or pleasure is necessary, however, they either make prudential value a latent property of the objective goods—which, we have seen, is implausible—or else make the attraction or pleasure a constitutive element of the objective goods—which, we have seen, suggests the objective goods are not the ultimate bearers of

prudential value. What we have seen here, then, is further proof of the claim that objective list accounts cannot plausibly account for our internalist intuition.

Let us take stock. We have considered a number of ways of formulating the plausible view that well-being consists in enjoying the good. One way of doing this begins with the thought that experiencing pleasure and possessing objective goods are each necessary conditions for an increase in well-being. This thought left us with three formulations of the view—one for each of the proposed constituents of a subject's welfare: experiencing pleasure, possessing objective goods, or both. Each of these formulations failed. But this left us with a more promising approach. This began with the thought, not that pleasure and objective goods are each necessary for well-being, but that the contribution one of these things makes to a subject's well-being may be enhanced by the presence of the other. By adopting this approach, we shifted our focus from the view that well-being consists in taking pleasure in the good to the view that well-being *in its fullest form* consists in this. One version of this latter view claimed that the extent to which certain objective goods contribute to one's welfare in part depends on the extent to which one takes pleasure in those goods. This version failed because it implied that there is some objective good that might make a life go better even if the life does not include a single experience of pleasure. The other version claimed that the extent to which experiences of pleasure contribute to one's welfare in part depends on the extent to which the objects of those pleasures are objectively valuable. This is the view we have called objective hedonism. We have not yet seen any reason to reject it.

As it happens, a theory quite like objective hedonism has been suggested by Fred Feldman.[25] Feldman ultimately rejects the view, but he offers it as an option for those who

---

[25] Feldman 2004 (ch. 5) and Feldman 2010 (ch. 10). The following comments concern the latter.

may find it more attractive than he does. As he sees it, we do not need such intricacy to explain what well-being consists in; indeed, he thinks well-being is simply a matter of the intensity and duration of our pleasurable experiences. Feldman argues for this in a roundabout way. Since he is primarily focused on the nature of happiness, Feldman argues that (1) happiness is simply a matter of the intensity and duration of our pleasurable experiences, that (2) the happy life is the life of welfare, and that (C) welfare is therefore simply a matter of the intensity and duration of our pleasurable experiences. While I agree with the first of these claims, I believe the second and third must be false. In what follows, I will argue that Feldman's claim that well-being consists in the experience of pleasure is indeed false. As I see it, we need something more akin to objective hedonism to fully explain the nature of well-being. Before that, though, I'll explain Feldman's account of happiness and argue for its plausibility. And this begins with a preliminary discussion of the nature of pleasure.

3 Pleasure, Desire, and Happiness

Hedonism has long been associated with the thought that what is good for us is the experience of *sensory pleasure*—or the type of pleasurable feeling or sensation we experience when our body is stimulated in certain ways. We might experience pleasure of this sort when we eat, drink, laugh, get a massage, and so on. The thought that all and only experiences of this sort are non-instrumentally good for us is often summarily rejected on the grounds that it is a "philosophy of swine"—equating our good to the good of non-rational animals like pigs.[26] This version of hedonism, it is said, neglects other features of human experience, and other things that we are uniquely positioned to attain, such as intellectual and moral goods. Understood in this way, hedonism seems untenable.

But this is, of course, not the only way of understanding hedonism, for it is not the only way of understanding pleasure. There is another type of pleasure that the hedonist might refer to when they claim that all and only experiences of pleasure are non-instrumentally good for us. This is *attitudinal pleasure*—or the type of pleasure that is essentially an attitude.[27] We experience pleasure of this sort when we take pleasure "in" things, such the fact that we have made progress on some project, or that our favorite team has won. Like sensory pleasure, attitudinal pleasure also has an opposite—attitudinal *dis*pleasure. We experience such displeasure when we are displeased or "pained by" some

[26] Mill 1863 (ch. 2).

[27] See, most recently, Feldman 2010 (pp. 109-118).

fact, such as the fact that our progress has stalled, or that our favorite team has lost. Attitudinal pleasures nevertheless differ from sensory pleasures in two main ways. First, attitudinal pleasures are always directed onto objects, which are generally states of affairs. In this way, attitudinal pleasures are like beliefs and hopes and fears—they are attitudes towards an object. Second, attitudinal pleasures need not be experienced in connection with any sensation. Though these pleasures may be associated with a certain phenomenology—namely, what it is like to be pleased or glad about some fact—we can experience them without experiencing any bodily feeling or sensation, just as we may have a belief or suspicion without experiencing any such feeling or sensation.

Before discussing any implications of this distinction for hedonism, it is worth briefly mentioning three other important and interesting features of attitudinal pleasures. First, attitudinal pleasure and displeasure come in degrees: a subject may take (dis)pleasure both in the fact that p and the fact that q, and yet take more (dis)pleasure in p than in q. It might be, for example, that a subject takes some pleasure in the fact that one of their wagers has yielded five dollars, and she takes a greater amount of pleasure in the fact that a separate wager has yielded ten dollars. And these pleasures may be outweighed by the displeasure she takes in the fact that a third wager has lost her one hundred dollars.

Second, the experiences of attitudinal pleasure that are relevant to happiness are experiences of *occurrent* attitudinal pleasure. These are experiences in which the object of one's pleasure is "on one's mind," or being thought about at that very time. Thus it is only when a subject is thinking about the state of affairs that she takes pleasure in that her experience will contribute to her happiness. This is to be contrasted with cases where a subject may appropriately be said to take pleasure in some state of affairs despite failing to think about that state of affairs at that very time. It might be, for example, that a subject very

much likes her job, which she has had for many years. Here, we might say: for as long as she's had that job, she's been pleased to have it. By this we do not mean: for every moment that she's had that job, she's taken pleasure in the fact that she has it at that moment. Rather, we mean: for as long as she's had the job, whenever she has considered the fact that she has it, she has taken pleasure in that fact. She is, in other words, *disposed* to take pleasure in the fact that she has the job—or *dispositionally pleased* about it. One could hold this attitude even when one was asleep, or focused on other things. Attitudinal pleasures of this sort do not constitute one's happiness. One can be happy at a time only if one is thinking about the things that make one happy. Accordingly, only experiences of occurrent attitudinal pleasure can contribute to one's happiness.

Third, occurrent attitudinal pleasure entails belief. In order to take occurrent attitudinal pleasure in some state of affairs, one must believe that that state of affairs obtains. A subject could not, for example, take pleasure in the fact that they were drinking coffee if they did not believe that they were drinking coffee. To take pleasure in that fact, the subject must believe that they are drinking coffee. Of course, this is not to say that it must be *true* that the subject is drinking coffee. Attitudinal pleasure does not entail true belief. A subject may take pleasure in some "fact" even if the relevant state of affairs fails to obtain—they may take pleasure in the "fact" that it is only two o'clock when in fact it is three. Still, the subject must believe that the state of affairs obtains—that it is two o'clock—to experience occurrent attitudinal pleasure.

How, at any rate, does this concept of attitudinal pleasure relate to hedonism? Well, if the hedonist claims that well-being consists in the experience of attitudinal pleasure, the objection that her view is a philosophy of swine seems to lose much of its force. This is because she can then say that the pleasures we take in reading great books, acting virtuously,

and the like, may be important components of our well-being. Whereas pigs cannot take pleasure in these things, we can. By focusing on attitudinal pleasures in defining what is good for us, the hedonist may thus account for such uniquely human goods. Furthermore, this hedonist can account for the contribution sensory pleasures may make to our well-being. After all, it seems plausible that whenever we experience a pleasurable sensation, we take pleasure in the fact that we are experiencing that sensation. That is, when we experience feelings of pleasure, we adopt the attitude of being pleased by the fact that we are experiencing those feelings. In that case, any increase in well-being brought on by an experience of sensory pleasure can be accounted for here by appeal to the corresponding experience of attitudinal pleasure. In fact, some have argued that the two types of pleasure do not just correspond or covary in this way, but attitudinal pleasure is actually constitutive of sensory pleasure. According to Feldman, in particular, sensory pleasure *just is* attitudinal pleasure taken in some feeling or sensation.[28] If that is right, the attitudinal hedonist can surely account for the contribution that experiences of sensory pleasure may seem to make to our well-being, since any experience of sensory pleasure is also an experience of attitudinal pleasure.

---

[28] This is one way of dealing with hedonism's problem of the "heterogeneity" of pleasure. The problem, roughly speaking, is that there appears to be no one positive feeling tone common to all the things that we enjoy or get pleasure from. The very dull pleasure we might get from reading, for example, seems to share little with the pleasure we might get from eating, and both of these seem to share little with the pleasure we might get from the accomplishment of one of our lifelong goals. Given this apparent heterogeneity, offering an account on which pleasure is defined in terms of its immediately recognizable feeling tone—or, what Sumner calls an "internalist" view of pleasure—will not seem a promising option (1996, p. 87). The alternative is to offer an account on which pleasure is defined in terms of some attitude, external to the feeling itself, that we take towards the feeling. By focusing on something outside of the feeling itself, however, such "externalist" views may seem to miss out on something important about the phenomenology of pleasurable experience. The heterogeneity problem is pointed out by Griffin (1986, p. 8), and discussed, with different results, by Crisp (2006, pp. 103-111), Feldman (2004, pp. 79-90), Sidgwick (1907, p. 125-127), and Sumner (1996, p. 92-93). Feldman's response evades this problem by appealing to the attitude we adopt to all pleasurable sensations: what is common to them is that we take attitudinal pleasure in them all. It is thus a version of externalism. I discuss where I stand on this issue at length in chapters 7 and 8.

A version of hedonism that takes the basic currency of non-instrumental goodness to be all and only experiences of attitudinal pleasure may thus seem far more plausible than one taking it to be all and only experiences of sensory pleasure. A view of this more promising sort has, in fact, recently been advanced and defended by Feldman. To reach the view in question, Feldman first argues that all and only experiences of attitudinal pleasure are the basic constituents of a person's *happiness*. This is surprising, given that happiness seems to differ from well-being in a number of important ways. It seems clear enough, at the very least, that the concepts are distinct: we are often more willing to attribute happiness to a subject than we are willing to attribute a high level of well-being to a subject. We may, for example, readily concede that a person who gets great pleasure from watching innocent people being tortured is happy, but we may just as quickly recoil at the thought that the person is doing well. This divergence in judgment at least implies that we do not use the two concepts in the same way, which suggests that the concepts are indeed distinct. Moreover, we can mark a fundamental difference between the concepts themselves. Whereas claims about a person's well-being, or about how well a person's life is going, are evaluative, claims about a person's happiness are merely descriptive.[29] When we say that a person is happy, we seem to simply describe the person's state. Yet when we say that a person is doing well, or that they are flourishing, we seem to both describe the person's state and offer a positive evaluation of that state. Similarly, when we say that a person has raised their left arm, we merely describe their act; yet when we say that the person has raised their left arm well, we both describe and evaluate their act. If we were to find that raising one's left arm was equivalent to raising one's left arm well, this would no doubt be a surprising result. It would

[29] As Feldman (2010, pp. 9-10), Haybron (2008, pp. 29-31), and Kraut (2007, pp. 132-133) observe.

be equally surprising to find that a life that has gone well just is a happy life. And that is what Feldman ultimately argues.

In support of his claim that the basic constituents of happiness are all and only experiences of attitudinal pleasure, Feldman considers a range of cases that cannot be accounted for by alternative accounts, but can be by his own. It seems that any case where a person is happy over some interval is a case where the person takes a greater amount of attitudinal pleasure in the propositions they are thinking about during that time than they take attitudinal displeasure in those propositions during that time. As Feldman sees it, this is because to be happy over some interval just is to take more attitudinal pleasure than displeasure in the propositions one thinks about over that interval. Competing accounts, on the other hand, such as "life-satisfaction" and sensory hedonist approaches, cannot account for this correlation. More specifically, accounts that take a subject's actual or hypothetical judgments of life satisfaction to determine their level of happiness cannot account for cases where a subject takes a great deal of attitudinal pleasure in the propositions they think about over the relevant interval yet the subject fails, or would fail, to report that they were happy over that interval. Similarly, accounts that take all and only a subject's experiences of sensory pleasure to determine their level of happiness cannot account for cases where a subject takes a great deal of attitudinal pleasure in the propositions they think about over the relevant interval yet the subject also experiences a surplus of sensory displeasure during that interval.

Suppose, for example, that a subject is experiencing intense sensory pain.[30] After being given an efficacious drug, the pain is significantly reduced, but not eradicated. Our subject still feels a very mild discomfort. Nonetheless, the subject takes great pleasure in the

[30] See Feldman 2010 (p. 33).

fact that the pain has been so significantly reduced. It is a big relief—she was beginning to wonder if the intense pain would linger forever. In this case, it seems plausible that as soon as the effects of the drug kicked in, the subject took enough attitudinal pleasure in the fact that the pain was relieved to outweigh the small amount of attitudinal displeasure she continues to take in the fact that she is experiencing (now-mild) sensory pain. That is, it seems plausible that as soon as the effects of the drug kicked in, her net balance of attitudinal pleasure and displeasure became positive. It also seems plausible that, at the same point, our subject became happy. Feldman's view can easily account for this intuition: over the course of the relevant interval, she takes more attitudinal pleasure in the fact that the pain is reduced than she takes attitudinal displeasure in the fact that she is in pain; thus, she is happy over that period. The sensory hedonist approach, on the other hand, cannot make sense of our intuition here. That's because our subject experiences more sensory pain than sensory pleasure over the relevant interval, and so must be unhappy, on this approach. If a case of this sort—where a person is happy over some interval despite experiencing more sensory pain than pleasure over that interval—is even possible, it follows that sensory hedonism about happiness is false.

Likewise, life satisfaction accounts have difficulty accounting for cases that the attitudinal hedonist about happiness can handle with ease. While there are, as Feldman notes, many possible versions of life satisfactionism, the accounts generally take a subject to be happy if and only if and because that subject is, or would under certain conditions be, satisfied with the conditions of their life as they perceive them. Yet we can easily imagine cases where a subject seems happy but never considers whether they are satisfied with their

life.[31] The subject may prefer to take life as it comes, to live in the moment. They never stop to consider whether their life is living up to their standards. Still, the subject may take a great deal of pleasure in all of the features of their life. What's more, the subject may be disposed to negatively evaluate their life if ever they were to actually consider it. And this may be the case, not because their life is not living up to their standards, but because evaluating their life would bring on a shift in their standards—no life would be good enough. In this case, if the subject never considered whether they were satisfied with their life, and they nevertheless took great pleasure in the features of their life, it seems plausible that their life might have been a happy one. According to our two forms of life satisfactionism, however, this plausible thought must be wrong. Since our subject never considers whether they're satisfied with their life, a view on which happiness consists in being satisfied with the conditions of one's own life implies that our subject cannot have been happy. And since our subject is disposed to be dissatisfied with their life on reflection, a view on which happiness consists in being satisfied with the conditions of one's life under reflective conditions has the same implication. Thus if we find it plausible that our subject here did indeed have a happy life, these life satisfactionist theories of happiness must also be rejected.

So, it seems that Feldman's view of happiness—or what we can call *attitudinal hedonism about happiness* (AHH)—has a substantial leg up on its primary competition. It gets us the right result in the greatest number of cases, and seemingly gets us the wrong result in none. Accordingly, I believe that we should accept Feldman's view of happiness, or else something very much like it. We should accept that all and only experiences of attitudinal (dis)pleasure contribute to a subject's (un)happiness; and that the extent to which an

[31] See Feldman 2010 (pp. 81-90).

experience of attitudinal (dis)pleasure contributes to a subject's (un)happiness is determined by the duration and intensity of that experience—the longer or more intense the experience, the greater increase in (un)happiness in the subject. That, I think, is the best way of accounting for happiness.

According to Feldman, however, it is also the best way of accounting for well-being. It seems clear that happiness tends to track welfare: when a subject is living a happy life, it seems likely that they are also doing well. Yet there are at least two types of case that seem to suggest that this connection is at best imprecise—or that not every very happy life must also be going very well. Cases of these two types have played a significant role in debates over welfare and happiness. In fact, we have already mentioned them: cases involving *deception*, and cases involving *immorality*. Many of us share the intuition that such cases disprove the claim that happiness precisely tracks welfare. But Feldman is not one of these people. He believes that we must give up the mistaken thought that these cases reveal a significant divide between happiness and well-being. To evaluate his arguments here, let us first consider a case of each of the two types.

3.1 Does happiness precisely track welfare?

Consider first cases of deception. These involve subjects who are very happy, whose lives seem to be going quite well from the "inside," but whose happiness is built on false beliefs. Imagine, for example, a deceived businessman.[32] This man is loved by his partner and family, is respected by his community, and has founded a successful business. That, at least, is what he thinks. In fact, however, the man's partner cheats on him, his family pretends to love him

---

[32] An example I owe to Kagan (1994, p. 311).

so that he will give them money, his community secretly despises him, and his business will soon fail thanks to a business partner who cooks the books. In this case, we begin with a picture of a subject who seems to be doing quite well. As soon as we realize that the subject is being deceived, however, the picture alters significantly. Now the man merely *seems* to be doing well—he merely *thinks* that he is doing well. Yet many of us will disagree with the man's judgment here. Sure, he *thinks* he's doing well, but he's not *actually* doing well. He's very *happy*, but he's not very *well off*.

Notice that the claim here is not that, while the subject is happy, his life in fact has *no* well-being. While it may seem plausible in this case that the subject's well-being is so radically reduced by deception, in other cases the deception may not be so pervasive. And, in these other cases, we might think that the subject's level of well-being is only slightly reduced. So we should not say that, in any case where a subject is deceived, and where this deception seems to negatively impact the subject's well-being, the subject's level of well-being must be zero. We may not even want to say that about the current case. But regardless of whether we think the deceived businessman's level of well-being is entirely or just partially diminished, the point is that it *has* diminished. Or, at least, that it seems to have.

On the other hand, the deceived businessman's happiness seems to have remained constant. Regardless of whether he is deceived, he is happy. In fact, he is *just as* happy. Since the deception does not affect his mental state, our description of his mental state does not change when we learn that he is deceived. And since claims about happiness are just descriptions of a subject's mental state, our claims about the subject's happiness here also do not change. Both before and after we learn of the deception, we claim that he is happy to degree x. Deception, accordingly, cannot in itself affect a subject's happiness.

Something similar seems to be true of immorality. To see this, consider two subjects. Each experiences the same amount of attitudinal pleasure and displeasure over the course of their life. And, on balance, each experiences much more attitudinal pleasure than displeasure—each lives a very happy life. The only difference between the subjects is that, whereas the pleasures of one are taken in unobjectionable things, such as the fact that one is loved by one's family and has been successful in one's aims, the pleasures of the other are taken in morally repugnant things, such as the fact that some innocent people are being tortured and that one gets to watch. That is, in every instance, the pleasures experienced by one have an unobjectionable object, and the pleasures experienced by the other have a morally repugnant object. In this case, many of us share the intuition that the subject who takes pleasure in morally unobjectionable things is better off than the subject who takes pleasure in morally repugnant things. Again, we need not claim that the latter subject has *no* well-being; we may say just that they are worse off than their morally unobjectionable counterpart to some extent. The point, however, is that one subject does indeed seem better off than other. And this suggests, as we have already seen, that the contribution a pleasure makes to one's welfare partly depends on the pleasure's object.

In contrast, it does seem plausible that the two subjects are equally happy. They each experience the same amount of attitudinal pleasure and displeasure. So, regardless of what they take pleasure and displeasure in, it is true that they are equally happy. Since the fact that the object of one's pleasure is (or seems to us to be) morally repugnant does not affect one's mental state, our descriptions of the subjects' mental states do not change when we learn that the objects of their pleasures are (or seem to us to be) morally repugnant. And since claims about happiness are just descriptions of the subjects' mental states, our claims here about the subjects' happiness do not change, either. Their mental states are such that each

takes the same amount of attitudinal pleasure and displeasure in things as the other. When we specify which things these are, this fact remains unaffected. As a result, it seems that each subject is equally happy.

So, again, we have a case where there seems to be a difference in welfare, but no corresponding difference in happiness. Since people like Feldman claim that happiness precisely tracks welfare, these cases present a substantial difficulty for their accounts. After all, if happiness precisely tracks welfare, we should not be able to find any case where the two come apart. Yet this, it seems, is just what we've found: cases where welfare varies but happiness remains constant.

Is there any plausible response available here? The results, I think, are mixed. While Feldman does not explicitly consider how his account might handle cases of deception, he does consider how it might handle cases of immorality. He considers a case where a person undoubtedly has a happy life, but the source of this person's happiness is in every instance (morally) "disgusting."[33] The person cheats, steals, kills, lies—all in the service of having a happy life. The person is then just like our subject who takes pleasure only in morally repugnant things. According to Feldman, we may be making the following sort of mistake when we appeal to these cases in attempt to show that happiness and welfare can come apart. Since the subjects in these cases are no doubt happy:

[33] Feldman 2010 (ch. 10).

1. If the claim that happiness precisely tracks welfare is true, then each of these subjects led a *welfare-good* life.
2. Each of these subjects did not lead a *morally good* life.
3. Therefore, the claim that happiness precisely tracks welfare is not true. (Feldman 2010, p. 207, slightly adapted)

Clearly, this reasoning is flawed: the fact that the subjects did not lead a morally good life does not imply that they did not lead a welfare-good life; and so, the conclusion does not follow.[34] We must be careful, then, not to mix up the notion of a morally good life with that of a welfare-good life. We are concerned with the features of a life that make it good, in itself, for the person who lives it. While these features may overlap with the features of a life that make it morally good, they also may not. So we must be careful not to tacitly switch from analyzing one of these sets of features to analyzing the other.

But perhaps we are not making that mistake. We might, on reflection, recognize this possibility of error, and nevertheless determine that our judgment is not about the moral deficiencies of the subjects but about their deficiencies in welfare. Suppose so. Still, Feldman thinks that our judgment is wrong. To point out what is wrong with it, he offers the following argument on our behalf. Again, since the immoral subjects are happy:

[34] Of course, it could be argued (by, for instance, Aristotelians) that failing to lead a morally good life implies failing to lead a welfare-good life, in which case the conclusion would follow. But that premise is not part of this argument.

1. If the claim that happiness precisely tracks welfare is true, then each of these subjects led a welfare-good life.
2. Each of these subjects did not lead a welfare-good life.
3. Therefore, the claim that happiness precisely tracks welfare is not true. (Feldman 2010, p. 207, slightly adapted)

This argument is valid; but, according to Feldman, it is not sound. The problem lies with the second premise. The claim that our immoral subjects did not lead welfare-good lives is untenable, Feldman thinks, because it implies that the subjects had no welfare whatever. This implication does not seem plausible, given the great amount of attitudinal pleasure the subjects experienced. It seems more plausible that the subjects had at least some welfare. In fact, it even seems more plausible that the subjects had a great deal of welfare. That is to say, the implication that our immoral subjects had a high level of well-being seems less implausible than the implication that they had no well-being. That being so, Feldman thinks, we should prefer the claim that happiness precisely tracks welfare to the claim that our subjects did not have welfare-good lives.

Notice, however, that we have not said that the subjects did not have welfare-good lives. Indeed, this point was stressed above: the claim that a subject who takes pleasure in morally repugnant things is worse off than a subject who takes pleasure of the same intensity and duration in morally unobjectionable things does not imply that the former subject has *no* well-being. Our claim was never that the immoral subjects did not lead welfare-good lives—or that the subjects had no welfare at all. Rather, our claim was that in virtue of the moral repugnance of the things they took pleasure in, their experiences of pleasure contributed to

their welfare at a reduced rate. Thus, our argument against Feldman is closer to the following. Once more, since our immoral subjects are happy:

1. If the claim that happiness precisely tracks welfare is true, then each of these subjects led a life that was *just as happy as it was well off*.
2. Each of these subjects did not lead a life that was just as happy as it was well off (the lives were happier than they were well off).
3. Therefore, the claim that happiness precisely tracks welfare is not true.

This argument properly expresses our objection to Feldman's view. If the fact that the immoral subjects took pleasure in morally repugnant things had any negative influence at all on their level of welfare—if it made them even slightly worse off than they were happy—it follows that Feldman's view is false. That's because he claims that well-being and happiness each consist in the experience of attitudinal pleasure; yet some attitudinal pleasures may contribute more to a subject's happiness than to their welfare. Since it seems that the moral status of the objects of our pleasures can affect our welfare, but cannot affect our happiness, we should reject the claim that happiness precisely tracks welfare. Happiness will fail to track welfare in any case where the objects of our pleasures are morally repugnant.

To this point, I have argued that attitudinal hedonism as a theory of happiness should be accepted, and that happiness, so conceived, does not precisely track welfare. It does not precisely track welfare because it is possible for a subject to experience a change in welfare without a corresponding change in happiness. This is evidenced, for one, by cases of immorality: a subject who takes pleasure in morally repugnant things seems to be doing worse, or to have a lower level of welfare, than a subject who takes pleasure of the same

intensity and duration in morally unobjectionable things; yet each seems equally happy. It may also be evidenced by cases of deception: a subject who takes pleasure in truths seems to be doing better than a subject who takes identical pleasure in falsehoods; but again, each seems equally happy. So while happiness tends to track welfare, it does not appear to do so in every case. If we are correct in our hypothesis that happiness consists in taking more attitudinal pleasure than displeasure in the propositions we think about over some interval, it follows that well-being cannot have that same source. Still, this may come as little surprise, given that claims about the amount of attitudinal pleasure and displeasure a subject experiences over some interval are merely descriptive, whereas claims about welfare are meant to be evaluative. Cases of immorality and deception demand an evaluation that claims about happiness alone cannot provide.

It seems, then, that there is something more to welfare than there is to happiness. And this something more seems to be related to *what* one takes pleasure in—or the *objects* of one's pleasures. Though Feldman rejects this intuitive line of thought, he recognizes its potential appeal. And so he offers a version of his view that can account for it. This view, I have said, is not unlike objective hedonism. In fact, it is a more precise version of objective hedonism—it further specifies what it is for the object of one's pleasure to possess "objective worth." On the view in question, some states of affairs *deserve* to have pleasure taken in them, and others *deserve* to have displeasure taken in them. Whether a state of affairs deserves to be the object of one of these attitudes depends both on features of the state—it may be one in which the world is at peace, for example, or one in which innocent people are being tortured—as well as on features of the subject who would take one of these attitudes towards the state. If, for example, a subject takes pleasure in a great work of art, but they have lied, cheated, and stolen their way to being able to see it, a state of affairs that would

otherwise deserve to have pleasure taken in it—the one in which the beauty of the art is seen—may instead deserve to have displeasure taken in it (in addition to shame, regret, etc.), solely in virtue of those features of the subject. In this way, we can determine the *desert-adjusted* amount of pleasure that a subject takes in an object: we factor in the pleasure's intensity and duration, and then adjust the pleasure's value in accordance with the deservingness of the pleasure's object. We can then say that well-being consists in the experience of pleasure, where the contribution an experience of pleasure makes to one's welfare is equivalent to the desert-adjusted amount of pleasure that experience contains. Following Feldman, we can call this view *desert-adjusted attitudinal hedonism* (DAAH).[35]

Clearly, DAAH is capable of getting us the right results in cases of immorality and deception. Whether it accomplishes this depends just on how we fill out the features of each case. So long as the desert-adjusted amount of attitudinal pleasure that our immoral and deceived subjects experience is slightly less than the desert-adjusted amount of attitudinal pleasure experienced by our morally unobjectionable and non-deceived subjects, the theory will tell us that the latter subjects are better off than the former subjects. Since it seems plausible that states of affairs that are morally repugnant, or that do not obtain, do not deserve to have pleasure taken in them—or, at least, as much pleasure as states of affairs that have the opposite features—it seems the theory will indeed get us the right results.

I have said that we should accept a theory that is similar to objective hedonism, and I have also said that DAAH is one such theory. Does this mean that I believe that we should accept DAAH? It does not.[36] On the contrary, I believe that we should accept one of two

---

[35] See Feldman 2010 (ch. 10). The view is introduced in Feldman 2004 (ch. 5).

[36] Here is one reason for doubting DAAH. It seems plausible enough that a subject might not deserve to take pleasure in some state of affairs, such as the fact that they are looking at a beautiful work of art, due to certain

other theories that are not unlike objective hedonism—theories that I will present shortly. But before that, I have made a further claim that stands in need of clarification. Specifically, I have said that we should accept attitudinal hedonism about happiness, or else something very much like it. In fact, I think that we should accept a theory of happiness that is somewhat wider in scope than AHH. To reach the view in question, it will be helpful to see that AHH is in fact equivalent to another theory of happiness—one that reduces attitudinal pleasure to more basic concepts. While this latter theory and AHH may amount to the same thing, it is not at all obvious that they do. So, in what follows, I will argue that this theory of happiness is equivalent to AHH. I will then express a worry about these theories. This analysis will lead naturally to the preferred theory of happiness, which will lead, in the following chapters, to the preferred theory of well-being. We begin, again, with a discussion of the nature of pleasure.

3.2 The nature of attitudinal pleasure

Recall that occurrent attitudinal pleasure requires belief. One natural response to this fact is to wonder what else it requires. Indeed, we might wonder whether attitudinal pleasure may be reducible to some combination of more basic attitudes. Obviously, attitudinal pleasure cannot be equivalent to belief alone: to believe something is not to take pleasure in it. To take pleasure in something is instead, at least in part, to have a "pro-attitude" toward it. It is,

---

features of the subject themselves, such as their having lied, cheated, and stolen their way to seeing the work of art. But what seems less plausible, I think, is that some states of affairs might deserve to have pleasure taken in them, quite apart from any morally relevant features of the subjects who would adopt this attitude towards them. The pleasure that, for example, a baby gets from playing with a toy, or that a person gets from taking a walk, does not seem deserved or undeserved, and nor do the relevant states of affairs seem to deserve or not deserve to have pleasure taken in them. Deservingness simply does not seem like the right normative concept to play this role. Now, it might be more or less *fitting*, or *appropriate*, to take pleasure in some states of affairs rather than others, but to say that the states of affairs are more or less *deserving* of having pleasure taken in them is, I think, to give an overly moralized account of our relationship to states of affairs.

in other words, to be "for," or "in favor of" it. Belief is plainly not an attitude of this sort. But given that attitudinal pleasure requires belief, and is also, in part, a kind of "pro-attitude," it might be that attitudinal pleasure just is a combination of belief and a pro-attitude of some sort. It is this possibility that I would like to explore.

Perhaps the most obvious suggestion for the relevant kind of pro-attitude is desire. On this picture, attitudinal pleasure is equivalent to belief and desire. Somewhat more precisely, to take occurrent attitudinal pleasure in some state of affairs is to occurrently both believe and desire that that state of affairs obtains.[37] So, for example, if a subject takes pleasure in the fact that they are drinking coffee, then they occurrently both believe and desire that they are drinking coffee. Similarly, if a subject occurrently wants to take a walk and occurrently believes that they are taking a walk, then they take pleasure in the fact that they are taking a walk. On the other hand, to take occurrent attitudinal displeasure in some state of affairs is to occurrently both believe that the state of affairs obtains, and desire that it does not (or, to be averse to it). Thus if, for example, a subject takes displeasure in the fact that they are drinking tea, then they occurrently believe that they are drinking tea and occurrently desire not to be drinking tea. Likewise, if a subject occurrently believes that they are waiting in line and occurrently desires not to be waiting in line, then they take attitudinal displeasure in the fact that they are waiting in line. More generally: a subject takes pleasure in something just in case they want it and believe they're getting it; and takes displeasure in something just in case they believe they're getting it and wish they weren't.

[37] It may be thought that this must be false, because desire is essentially prospective: one can only desire that p as long as one does not take p to be the case. Sumner (1996, pp. 128-130), for instance, argues that we can only have desires about the future. This is mistaken, however, for reasons Heathwood (2007, pp. 32-34) provides.

Has this suggestion any plausibility? According to Feldman, it does not. He offers two examples in support of this judgment. In the first, a subject, Nicko, has a desire of which he is ashamed.[38] Specifically, Nicko is ashamed of his desire to smoke his next cigarette. He is addicted, and is trying to quit. He knows that smoking his next cigarette will only make quitting harder. When he eventually gives in to temptation, he both believes that he is smoking the cigarette and desires to smoke it. Still, he does not take pleasure in the fact that he smokes it—on the contrary, he is ashamed.

Similarly, in Feldman's second case, a subject, Otto, has formed a desire to visit his elderly neighbor in the hospital.[39] He has only formed this desire, however, because he is a rigid Kantian moralist, and he believes that he has a duty of perfect obligation to visit his neighbor. He does not enjoy visiting this neighbor—and he certainly would not choose to do so, if it were not his duty—because the neighbor is ungrateful, unfriendly, and mean. When Otto eventually visits his neighbor, he desires to visit his neighbor and believes that he is visiting his neighbor, but he does not take pleasure in the fact that he is visiting his neighbor.

These cases, Feldman thinks, show us that we may have an occurrent belief and occurrent desire that p, without taking any corresponding occurrent attitudinal pleasure in the fact that p. And, if that is right, it seems to follow that occurrent attitudinal pleasure cannot be the combination of occurrent belief and desire. In fact, however, I believe that both of these inferences are mistaken. Notice first that we can secure the correct result in these cases without any alteration in the proposed analysis of attitudinal pleasure. All we

---

[38] Feldman 2010 (pp. 116-117).

[39] Feldman 2010 (p. 117).

must do, in each case, is appeal to a further desire that the subject presumably has. Given that he is ashamed of his desire to smoke his next cigarette, it seems plausible that Nicko must also have a desire not to have that desire—that is, a desire not to desire to smoke his next cigarette. In fact, we need not speculate—this is precisely what Feldman says: "He wishes he did not have that desire [to smoke his next cigarette]" (2010, pp. 116-117). This means that, when he lights up, in addition to his desire to smoke and his belief that he is smoking, Nicko will also have a desire *not* to smoke. And so, if his desire not to smoke is stronger than his desire to smoke, the current approach can easily account for Nicko's lack of attitudinal pleasure: he is taking more displeasure in the fact that he is smoking than he is taking pleasure in that fact. This is because, on this approach, attitudinal displeasure is just a belief that p and desire that not-p, and attitudinal pleasure is just a belief that p and a desire that p. So if a subject desires both p and not-p, whether they will experience attitudinal pleasure or displeasure when they believe that they are getting p will presumably depend just on which desire is stronger.

Similarly, we can account for Feldman's second case by appealing to the presumable fact that Otto has a further desire not to visit his neighbor, which he formed in virtue of his neighbor's unpleasant character traits. So long as this desire is stronger than his desire to visit his neighbor, the current approach tells us that Otto takes a greater amount of displeasure in the fact that he visits his neighbor than he takes pleasure in that fact. And so, this approach can account for our intuition that Otto does not, on balance, take pleasure in the fact that he visits his neighbor.

This response on behalf of the current approach leaves something to be desired, however. While plausible, it hinges on the plausibility of the claim that each subject's desire not to perform the relevant act is stronger than their desire to perform it. And given that

each subject ultimately chooses to perform the act, it may be objected that their desire to perform it must be stronger than their desire not to. On this suggestion, subjects always act on their strongest desire.[40] It is not possible for a subject to act on a desire that is weaker than another, competing desire: acting on a desire reveals that desire to be strongest. (Somewhat more precisely, on this picture, subjects always act on the strongest of their desires that they believe can be satisfied.) So—this objection goes—it cannot be the case that the subjects' strongest desires were the ones they failed to act on. And so the current analysis of attitudinal pleasure implies that each subject must have, on balance, taken pleasure in their respective states of affairs.

An alternative explanation here is that the problem instead lies with the current picture of motivation. After all, its implication that we cannot act on the weaker of some conflicting desires might simply seem not to accord with the phenomenology of everyday practical reasoning. We often feel that we act contrary to our strongest inclinations. It is rare, for example, that chores or errands are what we most want to do; those of us who most wish to work each day may be counted among the lucky; and more generally, few of us go through a day without experiencing what seems to be a clash between our strongest inclinations and our strongest reasons. While this may suggest that the proposed picture of motivation is flawed, it also may point to a deeper problem.[41] In particular, it may suggest that our current conception of desire is flawed, or is at least incomplete, contaminating any theory that makes use of that concept. Let me explain.

---

[40] This suggestion is inspired by a common (though ultimately misguided) conception of the "Humean Theory of Motivation," which claims that all motivation has its source in a relevant desire and means-end belief. Such a version of the theory is defended by Smith (1994, ch. 4); and attacked by Holton (2003).

[41] In fact, I do not think that it presents any problem for this picture of motivation, at least when the picture is adequately filled out. See Sinhababu 2009 and Sinhababu 2017 (ch. 2).

3.2.1 Two types of desire

There is an intuitive but somewhat elusive distinction between two separate senses of the word "desire." To desire in what is sometimes called the *narrow* sense is to have an attitude that governs your *affective* response to something.[42] This is also sometimes called "genuine" desire, because possession of this sort of desire involves an emotional or affective investment in the object of the desire. We form these desires when we are genuinely attracted or drawn to something, as when we consider an activity we enjoy or are passionate about. In contrast, to desire in what is sometimes called the *wide* sense is to have an attitude that governs your *actions*. This is also sometimes called desire in the "behavioral" or "dispositional" sense, since its possession involves a disposition to act (or, behave) in some way in certain conditions. When we have desires of this sort, we are disposed to act so as to bring the object of the desire about, or to make the desire's contents true. But that does not mean that we are genuinely attracted to the prospect of bringing that thing about.

This last point can be brought out by an example. When someone with whom we would like to spend our time is not genuinely attracted to the same things we are, we may correctly say: "I don't want to do what *I* want to do, I want to do what *you* want to do." We would be correct, in this case, if what we meant was that we desired in the behavioral sense to do whatever the other person was genuinely attracted to doing. If, for example, one is an ardent supporter of Liverpool Football Club, one will be genuinely attracted to the prospect of watching their next match. But if one's close friend or loved one finds soccer dull and

[42] The "narrow" versus "wide" distinction is made by Parfit (2011, p. 43), and my understanding of it has been helped by Hare (2013, p. 94). The same distinction has been made in different terms by Lewis (1988, p. 323), Nagel (1970, p. 29), Schapiro (2009, p. 230), and Sumner (1996, p. 120). For a more complete list of those who appeal to this distinction, see Heathwood 2019 (p. 674).

uninteresting, and is excited at the prospect of instead taking a walk, one may wish to give priority to this other person's genuine interest. Here, it would be appropriate for one to say that one desired, in the wide sense, to do what the other person desired, in the narrow sense, to do. One is not genuinely attracted to the prospect of taking a walk—or, at least, of missing the match—yet one is disposed to do just that.

This distinction helps us to see that there is a clear and intelligible sense in which one can do what one did not most want to do. The strength of a behavioral desire is determined by the hypothetical choices of the subject: one behavioral desire is stronger than another just in case the subject would choose to act on the one rather than the other. The strength of a genuine desire, on the other hand, is determined by the degree to which the subject is genuinely attracted to its object: in the most clear-cut cases, this will be a matter of the intensity (and presumably also the duration) of its affect—the greater its intensity and duration, the stronger the desire. Thus, whereas one cannot do what one does not most want, in the behavioral sense, to do, one clearly can do what one does not most want, in the affective sense, to do. That's because, whereas the definition of behavioral desire ensures that one always acts on one's strongest (behavioral) desire, it remains a possibility that the desire one acts on is affectively neutral, and that a desire one thereby spurns is affectively charged. In that case, one's strongest behavioral desire is not one's strongest affective desire; and, in that case, one acts on what is not one's strongest desire, in the affective sense.

3.3 The belief-desire theory of happiness

Returning now to Feldman's cases, it should be clear how this distinction can help us. I have said that Nicko's desire not to smoke his next cigarette is stronger than his desire to smoke it, and that Otto's desire not to visit his neighbor in the hospital is stronger than his desire to

visit his neighbor. According to the objector, this must be wrong, because Nicko ultimately smokes the cigarette, and Otto ultimately visits his neighbor, and it is not possible for a subject to do what they do not most want to do. We can now see where this objection goes wrong. In particular: while it is not possible for a subject to do what they do not most want, in the behavioral sense, to do, it is certainly possible for a subject to do what they do not most want, in the affective sense, to do. That being so, my claim that Nicko's desire not to smoke his next cigarette is stronger than his desire to smoke it, and that Otto's desire not to visit his neighbor in the hospital is stronger than his desire to visit his neighbor, need not be mistaken. It will not be mistaken so long as I am assessing the strength of their desires in the affective, rather than behavioral, sense. And that is indeed what I am doing. My claim is that Nicko's shame over smoking his next cigarette ensures that he is more genuinely attracted to the prospect of not smoking the next cigarette (or, more genuinely averse to smoking the next cigarette) than he is genuinely attracted to the prospect of smoking it. Nonetheless, since he ends up smoking it, he must have most wanted, in the behavioral sense, to smoke it. Likewise, Otto's aversion to his neighbor ensures that he is more genuinely attracted to the prospect of not visiting his neighbor than he is genuinely attracted to the prospect of visiting his neighbor. Yet, since he ends up visiting his neighbor, he must have most wanted, in the behavioral sense, to visit his neighbor. As we can see, my claims here do not conflict with the objector's. We agree that each subject does what he most wants, in the behavioral sense, to do. But I add, importantly, that neither does what he most wants, in the affective sense, to do.

This has an important implication for my hypothesis that attitudinal pleasure just is a combination of belief and desire. Since I suggest that Nicko and Otto are each, on balance, taking displeasure in what they are doing, and that whether they take pleasure or displeasure

in what they are doing depends on which of their desires is stronger, it follows that their desire not to do what they are doing must be stronger than their desire to do what they are doing. This, we have seen, can be true only if we are assessing their desires in the affective sense, since a subject cannot do what they do not most want, in the behavioral sense, to do. This means that, if I am right that attitudinal pleasure just is a combination of belief and desire, it must be that the desires in this combination are just genuine, or affective, desires. That is to say, it must be that attitudinal pleasure just is a combination of belief and *genuine* desire. For if attitudinal pleasure were a combination of belief and mere dispositional desire (or of belief and just any desire), it would not be possible for a subject's desire not to do what they were doing to be stronger than their desire to do what they were doing; and thus, it would not be possible for them to take displeasure in what they were doing. Since it is clearly possible for a subject to take displeasure in what they are doing, it cannot be dispositional desires that partly comprise attitudinal pleasures and displeasures. It seems, therefore, that genuine desires alone can play this role.

Does it seem right that attitudinal pleasure just is a combination of belief and genuine desire? I believe that it does. Each of the objections to this analysis that Feldman offers falls flat. This is because each of the objections he offers involves a subject who fails to take pleasure in something despite having a corresponding belief and dispositional desire for the thing. On the other hand, it is far harder to find an example in which a subject fails to take pleasure in something despite having a corresponding belief and genuine desire for the thing. In fact, as I see it, it is not possible to find such an example, because no such example could exist. Attitudinal pleasure just is a combination of belief and genuine desire: to take pleasure in something just is to believe that you are getting what you genuinely

desire.[43] More specifically: to take occurrent attitudinal pleasure in something just is to occurrently believe that you are getting what you occurrently genuinely desire; and to take occurrent attitudinal displeasure in something just is to occurrently believe that you are getting what you occurrently genuinely desire not to get.[44]

If that is right, then what does it mean for attitudinal hedonism about happiness? It means that this theory is equivalent to another theory of happiness, which we can call *the belief-desire theory of happiness* (BDH). This theory claims that all and only experiences in which a subject occurrently believes that they are getting what they occurrently genuinely desire contribute to the subject's happiness; that all and only experiences in which a subject occurrently believes that they are getting what they occurrently genuinely desire not to get contribute to the subject's unhappiness; and that the extent to which an experience in which a subject occurrently believes that they are getting what they occurrently genuinely desire (not) to get will contribute to the subject's (un)happiness is determined by the duration of the experience and the intensity of the desire—the longer the experience, or more intense the desire, the more (un)happiness that experience will contribute.

[43] And, in fact, I am not the only one who thinks this. Chris Heathwood (2006) has argued that attitudinal pleasure just is a combination of belief and desire. While he does not, in his 2006, distinguish between genuine desire and dispositional desire, he does make this distinction, and argue for its axiological relevance, in his 2019. It seems plausible, then, that Heathwood also accepts that attitudinal pleasure just is a combination of belief and genuine desire.

[44] Notice that this has an important implication for Feldman's proposed externalist understanding of sensory pleasure. If, as Feldman maintains, sensory pleasure just is attitudinal pleasure taken in a feeling or sensation, and, as I have argued, attitudinal pleasure just is occurrent belief that one is getting what one occurrently genuinely desires, it follows that sensory pleasure just is occurrent belief that one is experiencing some feeling or sensation that one occurrently genuinely desires. Is this view plausible? My arguments to this point seem to suggest so. More importantly, the view has firm empirical support. As two leading researchers in the field of the neuroscience of pleasure write: "Pleasure is never merely a sensation…Instead, it always requires the activity of hedonic brain systems to paint an additional 'hedonic gloss' onto a sensation to make it 'liked.'" (Berridge and Kringelbach 2008, p. 459). This "hedonic gloss," I suggest, is provided by genuine desire.

Since I have already argued for the plausibility of AHH, it follows from my arguments here that I have also already argued for the plausibility of BDH. So long as I am correct that occurrent attitudinal pleasure just is occurrent belief and occurrent genuine desire, it follows that AHH just is BDH. And so, any argument for one of these theories is also, it seems, an argument for the other. Accordingly, my arguments to this point in the chapter should lead us to see that the belief-desire theory of happiness is the most plausible of the presented options.

Admittedly, it still might seem that this theory faces decisive objections. This is the view taken, surprisingly enough, by Feldman himself. He offers a number of objections to a theory that is very much like—and, in fact, has served as the inspiration for—the belief-desire theory of happiness.[45] This is surprising because, if Feldman's arguments are successful, then he will seem to refute his own view: any objection to the belief-desire theory of happiness is also, we have seen, an objection to attitudinal hedonism about happiness. Fortunately for him, however, he needn't worry about this possibility, because none of his objections are successful. Seeing why they are unsuccessful will help us to secure our claim that BDH just is AHH.

3.4 Some unsuccessful objections

In attempt to refute the belief-desire theory of happiness—or, at least, something very much like it—Feldman offers three cases. In the first, a pessimistic graduate student, Susan, has

[45] The theory in question is presented in Davis 1981. As will be clear to anyone who is familiar with this incredibly stimulating piece, it has directly inspired many of the ideas in this chapter (and indeed, throughout this dissertation).

always been unhappy.[46] She is pessimistic about her future career prospects: she wants to get a job at a major research university, to publish in high-prestige journals, and so on, but she believes that the chances of this happening are very low. This is consistent with how Susan thinks about many of the features of her life: she often believes that she is not getting what she wants. Furthermore, she is often thinking about the fact that she is not getting what she wants—it is often on her mind. All of this implies that Susan is, on the belief-desire theory of happiness, unhappy. She thinks more about the perceived frustration of her desires than about their perceived satisfaction, and those of her desires that she believes to be frustrated are stronger than those that she believes to be satisfied. In light of this unhappiness, Susan goes to a counselor. After a number of failed attempts to help Susan, the counselor prescribes a drug, which is said to positively impact mood. Susan regularly takes the drug, and begins to feel cheerier, and more uplifted. She feels better. Still, few of her beliefs or desires have changed: she remains just as pessimistic about her career prospects and about the other features of her life as she was before taking the drug. So, Susan seems much happier, but her beliefs and desires have remained largely constant. If that is right, it suggests that one's happiness can change with no corresponding change in the perceived satisfaction of one's desires. And that suggests that the belief-desire theory of happiness is false.

Feldman's second case is similar to the first. There is a pessimistic graduate student, Glum, who shares all of the psychological features of Susan before she started taking the drug.[47] Glum is pessimistic and depressed. When Glum goes to see a counselor, however, the counselor's attempt to provide cognitive behavioral therapy is somewhat more successful

[46] Feldman 2010 (pp. 63-65).

[47] Feldman 2010 (pp. 65-66).

than it was with Susan. The counselor manages to get Glum to recognize and focus on his considerable academic talents. But in spite of this small success, Glum remains just as depressed as ever. He is no happier. If that is right, it suggests that the perceived satisfaction of one's desires can change with no corresponding change in happiness. And that suggests, again, that the belief-desire theory of happiness is false.

Feldman's third and final case is alleged to have the same implication as the second: that the perceived satisfaction of one's desires can change while one's happiness remains constant. In this case, a museum-goer, Lois, overhears someone talking while she looks at the museum's dinosaur exhibit.[48] The person remarks that it would be horrible to be eaten by one of these large and ferocious dinosaurs. Lois agrees. She certainly wants not to be eaten by a large and ferocious dinosaur. But, of course, Lois knows that dinosaurs are extinct, and that there is no chance of her being eaten by one. She is therefore unmoved by the fact that her desire not to be eaten by a dinosaur is satisfied: she is not pleased or glad about it, she just thinks it is obvious. So, Lois occurrently believes that she is getting what she occurrently wants, but it seems she is no better off for it. And so, once more, it seems we have a counterexample to the belief-desire theory of happiness.

Regarding the second and third cases, Feldman offers a diagnosis as to where the belief-desire theory goes wrong. He says: "happiness involves something more than, or something different from, merely believing things that you want to be true. As I see it, the missing element is more emotional or affective. In order to be happy, as I see it, one must be pleased about the good things that are happening" (Feldman 2010, p. 66). Since Lois and Glum's desires are missing this emotional or affective element, merely believing that they are

[48] Feldman 2010 (pp. 66-67).

satisfied is not enough to make either subject better off. But of course, on this point, we and Feldman can agree. If the subjects do not have genuine desires, and instead have mere dispositional desires, for the states of affairs to obtain, they cannot become any happier by believing that these desires are satisfied. The problem that Feldman points to in these cases is one that we have already solved. Lois and Glum are not made any happier by believing that their desires here are satisfied, because their desires here are merely dispositional—or, at least, that seems to be what Feldman is suggesting in his descriptions of the cases. They are not genuine desires. As a result, they lack the emotional or affective component that Feldman thinks necessary for happiness. Likewise, the belief-desire theory of happiness does not count the satisfaction of these desires as contributing to the subjects' happiness, because they are not genuine desires—they lack the relevant emotional or affective component. Accordingly, Feldman's arguments here do not threaten the belief-desire theory of happiness. They threaten only a version of this theory that fails to mark the distinction between genuine and dispositional desire.[49]

Turning now to the first case, we may seem to face a more difficult challenge. It seems to me, however, that the case is simply under-described. For one thing, we might question how psychologically realistic it is to assume that a person's mood can lift with almost no change to their beliefs and desires. Such a change in mood would presumably result in one's thinking that, for example, one was having a pleasant morning, or afternoon, or more generally that things were going well. These are all thoughts that Susan presumably did not have before she began taking the drug. And they are all thoughts that, on the belief-desire theory of happiness, might make a considerable contribution to Susan's happiness.

[49] Since Davis's (1981) theory does not recognize the distinction between genuine and dispositional desire, it is indeed subject to Feldman's objections here.

For another thing, we might wonder whether the drug caused any change in the things that Susan focused on in her thoughts. Even if her beliefs and desires remained largely constant, it might be that Susan began focusing more on those of her genuine desires that she believed to be satisfied, rather than on those of her genuine desires that she believed to be frustrated. This alone could presumably cause a large shift in one's mood—one is more often thinking cheery thoughts, and is thus cheerier—as well as one's happiness. The belief-desire theory tells us, again, that such a shift would indeed make Susan happier: she is thinking happy thoughts more often. Thus if, when we imagine the case, we are tacitly supposing that either of these two things occur to Susan after she begins taking the drug, our intuition that Susan becomes happier can be easily accounted for by the belief-desire theory of happiness. And if we are not assuming that either of these things occur, it seems far less plausible that Susan is any happier.

3.5 A worry

What I take all of this to show is that AHH is indeed no more than BDH. Any resistance to this claim seems to arise from a neglect of the fact that BDH concerns only our genuine desires. Once we recognize this fact, we secure the essential, affective component of happiness that desire and preference-based theories are sometimes said to lack. And this seems to secure the connection between BDH and AHH.

For all that, though, I worry that these theories may be false. My reason for worrying is that I think belief may just be too stringent a requirement for happiness. It seems plausible to me that a mere appearance, as opposed to a full-fledged belief, might be sufficient for enjoyment, and may thus be sufficient for happiness, when properly connected to genuine desire. That is, it seems plausible to me that enjoyment requires just that one sees oneself as

getting something that one genuinely desires. It then might not be that it requires that one believes that one is getting the thing. And, in that case, BDH and AHH are of course false.

Whether every instance of enjoyment entails a relevant belief on the part of the subject of the experience, or instead just a relevant appearance, turns on the nature of belief. When I see the road as wet, does it follow that I believe that I see it as wet? Must I consider whether it appears to me to be wet in order to conclude, and thus believe, that it is wet? Or is the appearance enough to entail an immediate, corresponding belief? I'm not sure. But here is why it matters: imagine that what I genuinely desire more than anything in the world is to see a wet-looking road. While driving along, you point off into the distance, and alert me to a road that appears wet up ahead. Upon seeing the wet-looking road, I'm overjoyed. You then inform me, despite knowing it will make no difference to my enjoyment, that the road is not actually wet, and that what I'm seeing is a mirage. Something to do with the hot air above the road. Since I believe you, I discard my belief that the road *is* wet, and come to believe that the road *is not* wet. But do I now also believe that the road *appears to be* wet? If not—if I still have nothing more than an appearance of the road as wet—then we have a case where mere appearance is sufficient for enjoyment. But if so, we would still have no reason to doubt that belief is necessary for enjoyment. Frankly, I'm unsure what to say about this case. Either way, the correct answer will turn on the nature of belief, and how belief relates to appearance. Since appearance is weaker than belief, the safe bet seems to be to require only appearance.

Consider also imagination. More specifically, consider the difference between, say, my genuinely desiring to throw a strike in some MLB game, and star pitcher Clayton Kershaw's genuine desire to throw a strike in that MLB game. Since throwing a strike is for a star pitcher a routine act, Kershaw believes that he will throw a strike in the game, and thus

enjoys the fact that he will throw a strike. In contrast, since I am not even in the city the game is in, I certainly do not believe that I will throw a strike in the game. Still, I can imagine what it would be like to throw a strike in the game, and I can enjoy imagining it. Here, it seems, I genuinely desire to throw a strike in the game, and have an imaginative appearance of myself throwing a strike, and thus enjoy seeing myself as throwing a strike. Again, the question is: is it necessary that I believe that I see myself as throwing a strike, and so have more than the mere appearance of myself as throwing the strike? If not, we again have a case suggesting that belief is not necessary for enjoyment, and that appearance is instead. But if so, we have a case giving us no reason to doubt the hypothesis that enjoyment requires belief. Unfortunately, I am again unsure what to say here, other than that the correct answer will turn on the nature of belief. Since appearance is weaker than belief, though—and since positing these extra beliefs about our appearances in each case seems to me somewhat unnecessary or unnatural—I will require only appearance in my theory of enjoyment, to avoid controversy.

3.6 The genuine desire theory of happiness

In that case, what is necessary for the experience of enjoyment? Only the ability to see oneself as getting what one genuinely desires. By applying this conception of enjoyment to Feldman's approach to happiness, we reach

> *The genuine desire theory of happiness*: all and only experiences in which a subject takes themselves to be getting what they occurrently genuinely desire are the basic constituents of the subject's happiness; all and only experiences in which a subject takes themselves to be getting what they occurrently genuinely desire not to get (that

> is, what they are occurrently genuinely averse to) are the basic constituents of the subject's unhappiness; and the extent to which an experience in which a subject takes themselves to be getting what they occurrently genuinely desire (not) to get will contribute to the subject's (un)happiness is determined by the duration of the experience and the intensity of the desire—the longer the experience, or more intense the desire, the more (un)happiness that experience will contribute.

This theory captures all that seems right about Feldman's approach, without implying that belief is necessary for enjoyment.[50] It thus gets us all of the right results in all of the cases considered here, and avoids a potential worry for the belief-desire theory of happiness. It seems, therefore, that the genuine desire theory of happiness is the most plausible theory of happiness available.

3.7 Taking stock

Here is where we are. I have argued that there is just one way of plausibly formulating the thesis that well-being, in its fullest form, consists in enjoying the good. On this formulation, well-being consists in the experience of pleasure, though the contribution an experience of pleasure makes to one's welfare in part depends on the extent to which the object of that pleasure is objectively valuable. We have called this approach objective hedonism. In attempt

---

[50] Though, of course, since appearance is weaker than belief, it also reaches the right result in cases where belief is necessary for the relevant kind of enjoyment. In some cases, for instance, we genuinely desire something in the future, such as that we will tonight read a mystery novel in a comfy bed. And it seems that we can presently enjoy the fact that this is how we will spend our night. To enjoy this "fact," we must take it to be a fact; which is to say, we must believe it. So, to enjoy some future state of affairs is just to genuinely desire that the state of affairs will obtain, and to take it to be the case that the state of affairs will obtain (or, to believe that it will). Such enjoyment is thus experience in which one takes oneself to be getting something that one genuinely desires.

to further explore this approach, we have considered Feldman's views of pleasure, happiness, and well-being. Though we've seen that important aspects of each of these views should be accepted, we have also seen that the views themselves ultimately must be rejected. Contrary to Feldman's suggestions, enjoyment does not require belief, and happiness does not precisely track welfare. These, however, are not the only point on which we should depart from Feldman. As we have seen, while we have good reason to accept as true aspects of Feldman's views of pleasure and happiness, we should reject his claim that these views are distinct from two others. More specifically, we have seen that attitudinal pleasure just is the combination of belief and genuine desire, and that attitudinal hedonism about happiness just is the belief-desire theory of happiness. While enjoyment does not require belief, and happiness therefore does not require attitudinal pleasure, we have seen that a wider view, the genuine desire theory of happiness, can capture the whole of the truth. This theory claims that happiness consists in enjoyment, which itself consists in taking oneself to be getting what one genuinely desires.

Here is where we are going. We now have the conceptual resources needed to construct a plausible theory of welfare. More precisely, we now have the resources needed to construct a plausible version of objective hedonism. Having seen the shortfalls of Feldman's claim that happiness precisely tracks welfare, we are left with a clear idea of what a theory of welfare must do, beyond ensuring the contingent connection between a subject's level of happiness and their level of welfare. Furthermore, we have the theory of happiness on which we can base this theory of welfare. Fulfilling this potential is the task of the following chapter.

4 The Genuine Desire Theory of Welfare

The genuine desire theory of happiness (GDTH), tells us that happiness is a matter of seeing ourselves as getting what we occurrently genuinely desire. The longer or more frequent our experiences of this sort, or the stronger our genuine desires in these experiences, the happier we are. That much, at least, has been established. What has not yet been established are the conditions under which we may become better off. More importantly, we have not yet established which things are non-instrumentally good for us, or what makes those things in particular non-instrumentally good for us. Our theory of welfare must offer some answer to these questions.

The objective hedonist approach tells us that a thing is non-instrumentally good for us if and only if and because it is an experience of enjoyment. That is, it tells us that all and only experiences of enjoyment are non-instrumentally good for us, and that anything that is non-instrumentally good for us is non-instrumentally good for us precisely because it is an experience of enjoyment. It also says that some experiences of enjoyment are better than others, not merely because they differ with respect to intensity or duration, but because the objects of the enjoyments have differing amounts of objective value. It leaves open, however, what is meant by both "enjoyment" and "objective value." Consequently, depending on our understanding of these concepts, we may each adopt this approach and nevertheless differ significantly in our views. Our primary task will therefore be to get clear on what we take these concepts to pick out.

Thankfully, we have already accomplished this for one of the concepts. We have already seen that enjoyable experience just is experience in which some genuine desire is taken to be satisfied. More precisely, we have seen that enjoyable experience just is experience in which we take ourselves to be getting something that we genuinely desire. We have also seen that taking pleasure in something—or, experiencing attitudinal pleasure—just is occurrently believing that one is getting what one occurrently genuinely desires. Of course, this may not seem to account for experiences of *sensory* pleasure, or pleasurable (bodily) *feelings*. But it also may be easily extended to do so. We can say that feelings of sensory pleasure are just sensations that we both genuinely desire and take ourselves to be experiencing. Sensations as diverse as those involved in taking a warm shower, riding a roller coaster, and eating tasty food are thus unified as pleasures in virtue of our genuinely desiring them when we take ourselves to be experiencing them.

Admittedly, it may be suggested that this picture does not fit with the phenomenology of experiencing sensory pleasure. More specifically, it may be thought that feelings of pleasure are in some sense more immediately felt, and not mediated by any attitudes that we adopt toward them. We seem to know of them immediately—we do not stop to form any desire about them before they begin feeling pleasurable. But even if we concede that much, still we can insist that whenever we experience these feelings, we must be thinking about them, and we must occurrently genuinely desire to experience them. That is, these are each necessary conditions for a felt experience to count as pleasurable. After all, if we are not thinking about a sensation—if it is not on our mind or on our conscious radar in any sense—it does not seem possible for us to be experiencing that sensation. A sensation that never enters our consciousness does not seem to count as a sensation that we experience. Thus, it appears, part of experiencing any sensation—including, of course,

sensations of pleasure—is representing oneself as doing so. Similarly, if we do not to any extent occurrently genuinely desire to be experiencing some feeling—if we are entirely affectively neutral or averse to experiencing it—it does not seem possible for that feeling to be a pleasure.[51] This seems clear enough when we are simply averse to an experience: if our affective stance toward the experience is one of pure aversion, the experience will plainly not seem pleasant. Likewise, if our affective stance toward some sensation that we are experiencing is truly neutral—if we have no affective preference at all as to whether it stays or goes—our experience does not seem to count as pleasurable.[52] (As it happens, it's not clear to me that it's possible to adopt such an affective stance (or lack thereof) toward a sensation. I cannot think of any case, normal or otherwise, where it seems plausible that a subject is truly affectively indifferent—and not merely affectively indifferent *on balance*—as to whether some sensation continues or ends. In any case, what does seem clear to me is that, if such cases are possible, the subject's experience of the sensation will not be pleasurable.) It seems, therefore, that both occurrent genuine desire, and occurrent thought that one is

---

[51] It might be objected: true enough, one cannot be affectively neutral or averse to a feeling that counts as a pleasure, but that does not mean that one must genuinely desire the feeling for it to count as a pleasure. Indeed, it might be that the affective component here is liking, rather than desiring. As Heathwood (2011, pp. 13-14) points out, however, liking requires belief, making it a compound attitude. Desiring, on the other hand, requires no other attitude. This suggests that desire is the more basic attitude, and the most natural suggestion as to what must be added to belief to reach liking is desire.

[52] Of course, it might be that one genuinely desires to be experiencing some feeling, and also genuinely desires, to the same extent, *not* to be experiencing that feeling. It might be, for example, that a subject has taken a pill that brings about feelings of pleasure just before going to a funeral for a loved one. When the pill kicks in, they might genuinely desire to be experiencing this feeling, simply because the feeling is of intense pleasure. Still, they might be aware of the fact that this is not an appropriate feeling to have at a funeral: they should be experiencing grief, not pleasure. As a result, they might genuinely desire not to be experiencing the feelings of pleasure, and might genuinely desire this to just the same extent that they genuinely desire *to* experience these feelings of pleasure. In this case, it might seem that although the subject's affective desires concerning the experience are neutral, the subject still manages to experience pleasure. Notice, however, that the subject's affective desires are merely *on balance* neutral—or neutral merely in virtue of being equal parts affective attraction and affective aversion. The subject still has a genuine desire to feel what they are feeling—their stance is not truly affectively neutral. And that is why their feeling seems to count as a pleasure. So, this is no counterexample to the claim that if we are entirely affectively neutral toward our experience of a feeling, our experience cannot count as pleasurable.

experiencing a sensation, are necessary for the experience of sensory pleasure. And so, regardless of the phenomenology of experiencing sensory pleasure, we must accept that these more basic psychological concepts are needed to explain the phenomenon of sensory pleasure. Those who would defend the view that pleasure is essentially a feeling are, accordingly, left with a claim that they must accept but cannot account for.

Furthermore, even if we concede that pleasure is essentially a feeling, or in any case that it is not essentially an appearance of getting what one occurrently genuinely desires, it does not follow that this latter psychological phenomenon is unimportant, or that it must itself be reducible to some positive feeling tone. It might be that it is simply distinct from, but often overlaps with, that of feeling pleasurable sensations. That is, even if pleasure is essentially a positive feeling tone, still it might be that there is some other, distinct kind of experience we have—one that we have come to associate with the term "pleasure"—where we take ourselves to be getting what we occurrently genuinely desire. In that case, we would just have to note that what we are concerned with here is only the latter kind of experience; our account would otherwise remain unchanged. Regardless of the nature of pleasure, then, the type of experience we are picking out here will retain any importance we may show it to have.

This much, at any rate, seems clear. Our version of objective hedonism will claim that something is non-instrumentally good for us if and only if and because it is an experience of taking ourselves to be getting something that we occurrently genuinely desire. It will also claim that the extent to which such experiences contribute to our welfare in part depends on which things we adopt this attitude towards—or what it is that we are enjoying. If we wish to stay true to the objective hedonist approach, our view will claim, more specifically, that the contribution such experiences make to our welfare will be directly

proportional to, not just their duration and intensity, but the objective worth of the things that we enjoy. That is: we are of course aware that, as a descendant of hedonism, this view claims that if our experience of enjoyment here is longer or more intense, it makes a greater contribution to our welfare; and if it is shorter or less intense, it makes a smaller contribution. But because our view is also a descendant of objective hedonism, it further claims that when the object of our enjoyment here has greater objective worth, this enjoyment makes a greater contribution to our well-being; and when the object has less objective worth, the enjoyment makes a smaller contribution.

So it seems we must now ask: in virtue of what, exactly, do the objects of our enjoyments have more or less objective worth? A complete answer to this question would appear to require a complete account of the nature of objective worth. And, as may be expected, I have no such account to offer here. Still, I think that we can go some way toward answering this question. The bulk of our understanding of which things have objective worth and of the extent to which they do is simply a matter of common sense. Few would disagree that a day spent marveling at impressive works of art would contain more objective worth than a day spent counting the blades of grass in one's yard. It seems clear, in addition, that those who would disagree with us here would be making an error by doing so: it is not that their tastes simply differ from, but are equally legitimate as, our own; rather, it's that they falsely believe that some activity has much more (or less) objective worth than it *actually* does. Objective worth, in other words, does not seem to be a mere matter of personal taste.[53] There is a fact of the matter as to which things have worth, in this objective sense, as well as the extent to which they do. Many of these facts are simply obvious—a matter of common

[53] Though I will offer a rather different analysis of our intuition here in chapter 8.

sense. That they are obvious itself becomes clear when we ask which of two things has more objective worth, or ask whether some thing has any. The answers, in these cases, present themselves to us immediately. In other cases, however, the answers are less clear.

It might be, for example, that a person is deciding between spending their day marveling at impressive works of art, or enjoying the company of a loved one. In this case, it is at least not obvious to me which of the two activities has more objective worth. Of course, it might be that the case is simply under-described: perhaps if we knew more about this person, their loved one, and the art, our answer here would be more immediate. But this, I think, need not be the case. As I see it, in some cases, the differences in the value of two things may simply be too slight for us to recognize, or may even themselves be imprecise.[54] Such cases reveal that objective worth is no obvious matter. Our comparisons of the objective worth of multiple things will not always admit of an obvious answer. Nor will our judgments about any single thing considered in isolation. Nonetheless, many of our comparisons and judgments here will seem clearly right or wrong. And this, I think, gives us a way of grounding more general claims about the nature of objective worth.

When we say that something has objective worth, what do we mean to express? Surely, by saying that the thing has "worth," we mean to express that the thing is in some way good or favorable. And by saying that this worth is "objective," we seem to mean that the thing has this worth in virtue of meeting some standard that is independent of any particular subject's stance toward the thing. That is, we are saying that the thing has a kind of worth that does not depend on any particular subject's conception of the thing's worth. This is to be contrasted with the type of worth a thing may have in virtue of a person's

[54] For more on this, see Parfit 2016 (pp. 113-115). See also Chang 2016.

circumstances and attitudes. A chair, for example, may be said to have a certain kind of value for a person in virtue of that person's legs being tired and their desiring to sit. Yet what the person thinks or feels about the chair cannot determine whether the chair has *objective* value. This point holds more generally: whether a thing has objective worth does not depend on what any particular subject thinks or feels about the thing. Putting these points about worth and objectivity together: as I see it, when we claim that something has objective worth, we mean to express that, quite apart from any particular subject's thoughts or feelings on the matter—including our own—the thing in question is in some way good or favorable.

More specifically, when we make claims of this sort, I believe that we are saying that the thing has certain features in virtue of which anyone capable of desiring the thing has reasons to desire it. That is, we are appealing to features of the thing that provide anyone capable of desiring it with reasons to desire it.[55] So, for example, when I say that the activity of looking at beautiful art has objective worth, I am appealing to the features of the activity—such as thinking about what the art depicts, what it represents, the use of colors, the brush strokes, and so on—that give anyone who is capable of desiring to engage in the activity sufficient reasons to desire to engage in the activity. I am, in effect, appealing to an independent standard by which anyone capable of desiring to engage in the activity can determine it to have this worth.

Why are these reasons available only to those capable of desiring to engage in the activity? For the same reason that facts about the world do not provide those incapable of

[55] This account is inspired by Mark Schroeder's (2010) discussion of the "right" and "wrong" kinds of reasons for desiring. The idea behind Schroeder's account, as I understand it, is that some activities give rise to a shared set of reasons—or a set of reasons that apply to anyone engaged in the activity. According to Schroeder, the "correct" way of engaging in the activity—and thus the desirable, or good way of engaging in it—is a matter of the extent to which that way of engaging in the activity is supported by the shared set of reasons. Thus, something is desirable for someone engaged in an activity just in case the set of reasons shared by anyone engaged in the activity tells in favor of that thing.

belief with reasons to believe those facts, and that facts about the consequences of φ-ing do not provide those incapable of intending to φ with reasons to intend to φ. The reasons just do not apply to those people. In the same way, facts about something cannot provide those incapable of desiring it with reasons to desire it.

This view has a number of intuitively plausible features and implications. Notice first that the extent to which a thing has objective worth, on this picture, can be easily accounted for in terms of the strength of our reasons to desire that thing. To say that something has more or less objective worth, then, is just to say that the reasons for anyone capable of desiring it to desire it are stronger or weaker. That is, it is just to say that the common pool of reasons for desiring the thing is more or less weighty. In this way, when we claim that one activity has more objective value than another, we are appealing to the reasons that any relevant subject has for desiring to engage in the one activity, and claiming that this set of reasons is weightier than the set of reasons that any relevant subject has for desiring to engage in the other activity. If we are right about the weight of these reasons, it seems to follow that any relevant subject who desires to engage in the other activity rather than the one, when the two conflict, is practically irrational. They are practically irrational because they've failed to desire that which they have most reason to desire, and have thus responded inappropriately to their reasons.

This also makes sense of the struggle one faces when trying to decide whether to spend one's day marveling at beautiful art or enjoying the company of a loved one. Each of these activities has many features in virtue of which anyone capable of desiring to engage in them has reasons to desire to engage in them. Indeed, without knowing more facts about the situation, it might be that there is no answer to the question which activity would have more objective worth, and thus best promote one's welfare. Moreover, it might be that the weight

of the reasons in favor of each activity are imprecisely equally weighty, in a way that makes their values in one sense incommensurable. Still, in other cases, the differences in objective worth between things will be much more obvious. It seems clear, for instance, that both marveling at beautiful art and enjoying the company of loved one have much more objective worth than watching innocent people being tortured. It also seems clear that watching innocent people being tortured has little, if any, objective worth. On the current picture, this is explained by the fact that there is very little about the activity of watching innocent people being tortured in virtue of which anyone has reasons to desire to watch innocent people being tortured.

Notice also that, on this picture, the reasons one has to desire something arise from the thing itself. It is in virtue of the thing's having certain features that our reasons to desire the thing arise. This has two important implications. The first is that, on this picture, we can make perfect sense of the fact that most all of the things that we have good reason to genuinely desire are things themselves, rather than the experiences to which those things are means. If, for example, we take virtuous activity to have great objective worth, this approach tells us that there are features of virtuous activity in virtue of which anyone capable of desiring to act virtuously has reasons to desire to act virtuously. If we form this desire, and go on to act virtuously, we may take pleasure in the fact that we are acting virtuously. But the thing that we have reason to desire remains virtuous activity, rather than the experience of pleasure that such activity might occasion. So, this picture predicts that welfare is a by-product of desiring things that possess objective worth. And that is precisely how it seems to us from the first-person perspective.

The second implication of this picture is that the reasons arising from many of the things we see objective worth in will not be available to those incapable of desiring those

things. So, since babies and cognitively advanced non-human animals are incapable of desiring many of these things, such as to engage in the activity of appreciating beautiful art, it will be no strike against these beings that they do not desire these things. That is, this picture predicts that the welfare of these beings will not be assessed by reference to a standard that they are incapable of meeting. These beings are not practically irrational for failing to desire to appreciate beautiful art, because they are incapable of such desire. But, it is also precisely because they incapable of such desire that their potential for well-being is not as great as ours. Such beings are not (yet) capable of having lives with as much welfare as us because they are not (yet) capable of desiring as many objectively valuable things, which suggests that they will have far less experience in which they enjoy things with objective value than we will. Again, this fits perfectly well with intuition.

Finally, notice what this picture implies about our claims about welfare. Recall that while claims about subjects' happiness are merely descriptive, claims about subjects' welfare are evaluative. That, on the current picture, is because the source of one's happiness is irrelevant to one's level of happiness. One could be happy for all of the wrong reasons—or, indeed, for no reasons at all. So long as one thinks one is getting what one genuinely desires, on the genuine desire theory of happiness, one is to that extent happy. Accordingly, claims about happiness aim only to describe the subject's mental state, specifically as it relates to the extent to which they're taking themselves to be getting what they genuinely desire. Claims about welfare, on the other hand, effectively aim to evaluate the subject's happiness. More specifically, on the current picture, claims about welfare are evaluations of the subject's genuine desires. To say that someone is well off is just to say that they are getting things that they genuinely desire, and that the reasons for anyone (capable of genuinely desiring those things) to genuinely desire those things are sufficiently weighty. It is to appeal to an

independent evaluative standard, accessible to anyone capable of genuinely desiring the things, that tells us the weight of our reasons to desire the things. In this way, when we say that a subject who takes great pleasure exclusively in watching innocent people being tortured is not doing well, we are appealing to the fact that this person is failing to genuinely desire what they have most reason to genuinely desire. There are many features of human suffering that give us reasons to be genuinely averse to it, and this subject is not responding appropriately to these reasons. As a result, they receive from us a negative evaluation. They are not genuinely desiring in accordance with the common pool of reasons. And they are, to that extent, irrational. This account of objective worth thus explains why it is that claims about welfare are evaluative, while claims about happiness are merely descriptive.

We are now in position to put all of this together. On the genuine desire theory of happiness, happiness consists in the experience of taking oneself to be getting what one occurrently genuinely desires. Welfare, on the other hand, must involve more than mere happiness. Specifically, it must involve a kind of objective worth. This, we have seen, is the basis of the objective hedonist approach. This approach claims that welfare consists in the experience of enjoyment, though the contribution such an experience makes to one's welfare in part depends on the extent to which the object of that enjoyment is objectively valuable. If we apply to this approach our analyses of enjoyment and objective worth, we reach

> *The genuine desire theory of well-being (GDTW)*: all and only experiences in which a subject takes themselves to be getting what they occurrently genuinely desire directly contribute to the subject's well-being; all and only experiences in which a subject takes themselves to be getting what they occurrently genuinely desire not to get (that is, what they are occurrently genuinely averse to) directly contribute to the subject's

> ill-being; and the extent to which an experience in which a subject takes themselves to be getting what they occurrently genuinely desire (not) to get directly contributes to the subject's well-being (ill-being) is determined by the duration of the experience, the intensity of the desire, and the extent to which the object of the desire admits of reasons for anyone capable of desiring (not to get) it to desire (not to get) it—the longer the experience, more intense the desire, or more such reasons for desiring (not to get) the desire's object, the greater contribution that experience makes to one's well-being (ill-being).

As a version of objective hedonism, this theory is both hybrid and monistic. And since it determines the contribution that an experience of the relevant sort makes to one's welfare by appealing to an independent evaluative standard—in particular, whether and to what extent the pleasure's object admits of reasons for anyone capable of desiring it to desire it—it seems the theory is objectivist.

Return now to the cases that disproved Feldman's claim that happiness precisely tracks welfare. These were, most notably, cases of immorality. It should be easy to see how the genuine desire approach can account for these cases. Since happiness consists in the experience of taking oneself to be getting what one occurrently genuinely desires, it seems that two subjects whose experiences of this sort are identical in terms of duration and intensity of desire, yet distinct in terms of the desire's objects, must always be equally happy. What the subjects genuinely desire does not seem relevant to their level of happiness. In contrast, it does seem relevant to their level of welfare. When one subject's genuine desires here are for morally repugnant things, and the other's genuine desires are for morally unobjectionable things, the former seems worse off than the latter in virtue of that fact. On

the current picture, this difference between happiness and welfare may be explained by the fact that one subject genuinely desires things that have a great deal of objective worth, and the other does not. This fact is irrelevant to their level of happiness, and so the two are equally happy. Yet it is certainly relevant to their level of welfare. When a subject's genuine desires are for things that there are very few reasons for anyone to desire—such as the needless torture of innocent people—their experience of taking themselves to be getting what they genuinely desire is worth far less, on the current account, than it would be if their genuine desires were for things that there are many reasons for anyone to desire. That being so, this account implies that the subject whose genuine desires are for morally repugnant things has a much lower level of welfare than their morally unobjectionable counterpart, even though the two are equally happy. The genuine desire approach thus secures the correct result in cases of immorality.

Consider next cases of deception. These cases, too, seemed to cast doubt on Feldman's claim that happiness precisely tracks welfare. Again, since happiness consists in the experience of seeing one's genuine desires as satisfied, it seems that two subjects whose experiences are identical, but whose appearances diverge in terms of veridicality, must always be equally happy. Whether the subject's appearance of themselves as getting what they occurrently genuinely desire is veridical or not does not seem relevant to their level of happiness. Yet it might seem relevant to their level of welfare. After all, it might seem that when one subject has a veridical appearance of themselves as getting what they occurrently genuinely desire, and the other—who has just the same experiences—has a non-veridical appearance, the latter seems worse off in virtue of that fact. If that is right, it appears to follow that the veridicality of such appearances can affect a subject's welfare without affecting their happiness. The genuine desire approach surely secures the latter result—it

implies that the veridicality of such appearances cannot affect a subject's happiness—but it admittedly does not seem to secure the former. That is, this approach does not seem to imply that the veridicality of these appearances can affect a subject's welfare. After all, the genuine desire theory of well-being says nothing to suggest that the contribution an experience of the relevant sort makes to one's welfare in part depends on whether one's appearance in the experience is veridical or not. The only such determinants are the duration of the experience, the intensity of the desire, and the objective worth of the desire's object. Accordingly, it does not seem to matter, on this approach, whether a subject is actually getting what they take themselves to be getting: their level of welfare is the same either way. It seems, then, that the genuine desire approach cannot secure the correct result in cases of deception, since it appears to imply that the mere fact that a subject is being deceived—or otherwise has a non-veridical experience of themselves as getting what they genuinely want—cannot affect the subject's level of welfare.

Contrary to appearances, however, the theory does not have that implication. It avoids this problem in the same way that it avoids the problem of immorality: by appeal to objective worth. Notice that, on the current approach, the objective worth of a given thing is a matter of the reasons for anyone capable of desiring the thing to desire the thing. And, in cases of deception, there are surely fewer reasons for anyone to desire the relevant objects of the deceived subject's desires than there would be if the subject were not deceived. That is, there are surely fewer reasons for anyone to desire the specific things that the deceived subject desires (such as that their specific business partner respects them), and this is precisely because the subject is deceived. We can no longer appeal to the fact that the subject's love, loyalty, and respect are reciprocated in grounding our claim that the objects of the subject's desires are worthy of being desired. As a result, the objects of the deceived

subject's desires have less objective worth than they would have if the subject had the same experiences but was not deceived.[56] And so, since the genuine desire theory of well-being says that the contribution an experience of enjoyment makes to one's welfare in part depends on the objective worth of the relevant genuine desire's object, the theory secures the correct result: the deceived subject is worse off than the experientially identical non-deceived subject.

It seems, then, that the genuine desire theory of well-being secures the correct result in the two types of case that led us to reject Feldman's claim that happiness precisely tracks welfare. The genuine desire theory of happiness also secures the correct result in these cases. Accordingly, if these cases are to be any judge, it seems the genuine desire approach correctly links happiness to welfare.

There is another important connection that the genuine desire theory of well-being can account for. This is the connection between each of the three classic types of well-being theory. Earlier, I suggested that these three types of theory each contain a kernel of truth, or some feature of the nature of well-being that they are accurately capturing. If that is correct, it could help to explain why it is that theories of these types in particular have proven so prominent: since each captures an important fact about the nature of well-being, a theory of any of these types will be likely to approximate the truth. It could also help to explain the

[56] This is, however, not to say that they have no objective worth, or that their objective worth is then fixed at this lesser amount. It is not that, for instance, the cheating partner's love, which the deceived subject desires, must now be worth less regardless of circumstance. Indeed, we could stipulate that the deceived subject has been abusive and thus driven their partner to cheat, or that their partner later learned from their infidelity and this helped them to have a healthier relationship in the future. In either case, the objective worth of their partner's love would seem to improve, and this is because there would be more reasons for anyone capable of desiring the love to desire it. On the other hand, their partner might continue to cheat and show no remorse. In this case, the objective worth of their partner's love would seem to stay at a diminished level, or even to diminish further, and this is because there would be equal or fewer reasons for anyone capable of desiring the love to desire it.

apparent stalemate in the current literature on well-being: each theory approximates the truth, but no one of them seems to capture the whole of it. The genuine desire theory of well-being accounts for these three kernels of truth, and thereby solves the problem that led us to this impasse.

The attraction of hedonism seems to lie in its claim that enjoyable experiences make us better off. A life with no such experiences does not seem like a good one, and any such experience seems to improve our lives to at least some degree. Similarly, a life full of painful experiences seems like a bad one, and any experience of this sort seems to make us at least somewhat worse off. The genuine desire theory of well-being implies that the kernel of truth in hedonism is, at least in part, indeed what made it attractive to us in the first place: the claim that all enjoyable experiences are good for us and all painful experiences are bad for us. As we have seen, however, this claim does not go far enough. We must add to it the further claim that only enjoyable experiences are good for us, and only painful experiences are bad for us. If something else were good for us, it would presumably be good for us all on its own; in which case, it would be good for us even in the absence of any enjoyable experience. Yet nothing seems good for us in the absence of such experience. Thus the genuine desire theory of well-being correctly implies that the kernel of truth in hedonism is its claim that all and only enjoyable experience is basically good for us, and all and only painful experience is basically bad for us. That it has this implication seems clear: it claims that all and only experiences of seeing ourselves as getting what we occurrently genuinely desire are basically good for us and that all and only experiences of seeing ourselves as getting what we occurrently genuinely desire not to get are basically bad for us. And so, since the former experiences are just experiences of enjoyment, and the latter are just experiences of

unpleasantness, the theory implies that all and only experiences of enjoyment are basically good for us, and all and only experiences of unpleasantness are basically bad for us.

Given that the genuine desire theory of well-being is a version of objective hedonism, it should come as little surprise that it can account for the kernel of truth in hedonism. It should be unsurprising as well, for that same reason, that it can account for the kernel of truth in the objective list approach. Our attraction to the latter seemed to owe to its objectivism—or its appeal to an independent evaluative standard in determining what is good for us. Yet our attraction here was tempered by the fact that appeals of this sort seem incapable of ensuring a proper connection between our attitudes and the things that are good for us. The genuine desire theory of well-being avoids this tension by appealing to an independent evaluative standard only in determining *how* good these things are for us. This leaves the theory free to claim that the things that are in fact good for us have that status in virtue of some fact that properly links the things to our attitudes towards them. In any case, the point here is that the theory captures the kernel of truth in the objective list approach, which is that the contribution the things that are good for us make to our well-being in part depends on the extent to which those things meet some independent evaluative standard. Given that the theory claims that the contribution an experience in which we see ourselves as getting what we occurrently genuinely desire makes to our well-being in part depends on the extent to which the object of our desire is objectively valuable, it again seems clear that it captures this kernel of truth.

What attracted us to desire satisfactionism, finally, was the importance it attributes to the link between what a subject desires and what is good for that subject. It seems implausible that what is good for a person might fail to link in any way with what the person desires or is attracted to. And while a person may fail to be attracted to certain experiences

or goods, it seems they cannot fail to be attracted to the objects of their desires. This, we have seen, is true only of genuine desires: people are often not attracted to the objects of their dispositional desires, but they must be at least somewhat attracted to the objects of their genuine desires. The genuine desire theory of well-being tells us that it is not the objects of these genuine desires that make us better off, nor is it their satisfaction; rather, it is the experience of seeing them as satisfied. In that case, however, it might seem that the genuine desire theory of well-being fails to capture the kernel of truth in desire satisfactionism. After all, while the proposed theory clearly makes some of our desires, as well as their perceived satisfaction, important aspects of our well-being—and these may be seen as further kernels of truth in desire satisfactionism—it nevertheless claims that what ultimately constitutes our well-being is a certain kind of experience. And since it seems that we can have these experiences without being attracted to or genuinely desiring *them*, the theory appears not to properly link what is good for us to what we genuinely desire or care about. If the need for this proper connection is the kernel of truth in desire satisfactionism, the genuine desire theory of well-being may seem not to capture it.

But again, appearances here are misleading. In truth, the proper connection between what is basically good for us and what we genuinely desire or care about is, for the most part, indirect. To be sure, many of the things that seem good for us are also things that we genuinely desire. We are often genuinely attracted to, for instance, health, autonomy, knowledge, love, and so on. Yet, while these things are often good for us, they are not *basically* good for us. They are instead good for us in virtue of the experiences they lead us to. Now, this may lead us to wonder: in that case, why do we genuinely desire these things, rather than the experiences themselves? That is, if what is basically good for us is the sort of experience these things lead to, why should it be that we are genuinely attracted to these

things, rather than the experiences they lead to? Here, though, we have a clear answer: a genuine attraction to these things, which are only derivatively good for us, allows us to experience a much greater amount of enjoyment than a genuine attraction to just enjoyable experiences would allow us to experience. If one genuinely desired only enjoyable experiences, one would be incapable of experiencing the enjoyment associated with, for instance, a loving relationship, because a relationship of that sort requires a genuine desire for (among other things) another person and their good. It is not that we just happen to tend to genuinely desire things that, like loving relationships, lead to enjoyment; it's that we tend to genuinely desire these things *because* they lead to enjoyment. This helps to show that the genuine desire theory of well-being in fact secures the proper connection between our genuine desires and our good. This connection is indirect. What is basically good for us is not determined by our genuine desires, but any realization of what is basically good for us will be mediated by our genuine desires. We reach the good by way of our genuine desires, though the good is not fixed by the contents of these desires. The genuine desire theory of well-being thus properly connects our genuine desires to our good.

In fact, as I see it, desire satisfactionism's most crucial flaw is in misidentifying this connection between desire and the good. Far from being the theory's kernel of truth, the thought that something can be basically good for one only if one is properly attracted to it ignores the fact that it is the experience of seeing oneself as getting things that one is properly attracted to that constitutes the real payoff. Still, something close to this thought remains true. This is the internalist intuition that something cannot be intrinsically valuable for one without a proper connection to what one finds attractive. As we have seen, this principle is true, though the connection it refers to is ultimately not a direct one. What is intrinsically valuable for one is not fixed by what one genuinely desires, but one's genuine

desires will always mediate instances of what is intrinsically valuable for one. The genuine desire theory of well-being thus accords with the internalist intuition, which is plausibly thought of as the kernel of truth motivating desire satisfactionism.

It seems, therefore, that the genuine desire theory of well-being meets all of the criteria for an adequate theory of well-being. It begins with the plausible thought that well-being consists in enjoying the good; and, in this case, this thought does not lead to any implausibility. Instead the view claims that all and only enjoyable experiences—or experiences in which we see ourselves as getting what we occurrently genuinely desire—are basically good for us, that all and only unenjoyable experiences are basically bad for us, and that the contribution such experiences make to our well-being (ill-being) in part depends on the objective worth of what we thereby enjoy (dislike). The theory thus avoids the problems that plagued alternative versions of the view that well-being consists in enjoying the good. Moreover, this theory is able to account for the kernel of truth in each of the three classic types of well-being theory. Like hedonism, it claims that all and only enjoyable experiences are basically good for us, and all and only unenjoyable experiences are basically bad for us; like the objective list approach, it claims that the contribution a thing makes to our well-being in part depends on the extent to which the thing meets some independent evaluative standard; and like desire satisfactionism, it implies that nothing can be good for us if it is not properly connected to what we genuinely desire.

This is, I believe, more than we can say about any alternative theory of welfare. The genuine desire theory of well-being can meet more of our demands than any other theory of well-being that we've considered here. Still, doubts may persist. In the following chapter, I'll consider some of these doubts, and argue that proponents of the genuine desire theory of well-being ultimately have no reason to worry.

5 Some Objections

The genuine desire theory of well-being is, I have said, the most plausible theory of well-being advanced thus far. One important reason for thinking this is that the genuine desire theory of well-being alone seems capable of capturing the kernel of truth in each of the three classic types of well-being theory. It makes sense of the debate up to this point, implying that each of these approaches has something going for it—something ensuring that specific formulations of these theories tend to approximate the truth—yet each fails to capture the entirety of the truth. This, however, may seem just as much a burden as it seems a blessing for the proposed theory. After all, by capturing the kernel of truth in each of the three classic approaches, this theory seems to open itself up to three corresponding sets of objections. Those who would praise the view for capturing the kernel of truth in one of these approaches would also seem likely to criticize it for capturing some element of a competing approach. It seems, accordingly, that the genuine desire theory of well-being will face attack from all angles. The aim of this chapter is to defend the theory from each such attack. I begin with some objections sometimes faced by hedonists, and then move on to others sometimes faced by desire satisfactionists. I return to cases of deception and false belief, which I have already discussed at length in the previous chapter, in chapter 7.

5.1 Joke book

On the genuine desire theory of well-being, all and only experiences in which one takes oneself to be getting what one occurrently genuinely desires are basically good for one, and

all and only experiences in which one takes oneself to be getting what one is occurrently genuinely averse to are basically bad for one. What one is thinking about—or what one occurrently sees oneself as getting—then seems to play an important role on the theory. After all, it seems plausible that, at any one moment, there is some genuine desire of ours that is satisfied, and some genuine desire of ours that is not satisfied. And so, since the experiences that determine our level of well-being are ones in which we see ourselves as getting what we genuinely desire or are genuinely averse to, it seems that, at any one moment, we could become better or worse off merely by thinking about one of these genuine desires rather than one of these genuine aversions. It seems, that is, that our level of well-being is determined by whatever we happen to focus on. If we focus on those of our genuine desires that appear to us to be satisfied, we are thereby made better off. If we instead focus on those of our genuine aversions that we take to be satisfied, we are thereby made worse off. So, given that, at any time, we have some genuine desire that it is possible for us to see as satisfied, and some genuine aversion that it is possible for us to see as satisfied, we seem, in effect, to have a choice about whether to improve or diminish our well-being at any given time. To improve it, all we must do is focus on those of our genuine desires that we take to be satisfied at that time; to diminish it, all we must do is focus on those of our genuine aversions that we take to be satisfied at that time. But that seems easy. Why, then, does becoming better off, and making others better off, normally seem much more difficult?

Let me try to put this challenge in a slightly different, slightly more forceful way. Consider a victim of grave injustice. This person lives in a third-world country. She has very little money, food, or education, and very few ways of acquiring these things. Furthermore, as a woman in this country, she has very few rights or liberties. Still, she is able to work,

doing back-breaking manual labor each day, giving her just enough to survive. An objector might point to this type of person and say: "the genuine desire theory of well-being implies that, to become better off, this woman should simply focus on those of her genuine desires that she takes to be satisfied. But that seems absurd. Surely this person simply lacks certain fundamental bases of well-being, such as autonomy, self-respect, education, and the like. We would seem to do much better by providing her with these things than we would by merely directing her to count her blessings. Yet the genuine desire theory of well-being does not reach this conclusion. The theory implies that, rather than ensuring that a person has goods such as health, autonomy, education, and the like, we could just as easily improve their level of welfare by giving them a book full of jokes, and telling them to focus. If they just think about the jokes, and ignore all of the more important things that they lack, their level of welfare could improve just the same. This conclusion is implausible. It seems clear that we should instead be ensuring that the person has these more important things. So, we should reject the genuine desire theory of well-being."

As I see it, the problem here is overstated. To be sure, the genuine desire theory of well-being implies that what we are thinking about over some interval determines our level of well-being over that interval. If we are thinking more about those of our genuine desires that we take to be satisfied than about those of our genuine aversions that we take to be satisfied, we will over that interval be doing well. If our thinking has the opposite structure, we will over that interval be doing poorly. That implication is, I think, perfectly plausible.[57]

[57] If you do not, consider: a person can seem to have everything going for them, yet suffer depression nonetheless; and a person can seem to have very little going for them, yet be constantly satisfied nonetheless. Regardless of how many of the things that we genuinely desire we are able to see ourselves as getting, then, our happiness, and thus our well-being, will depend in large part on our ability to focus on those things we do genuinely desire and see ourselves as getting, instead of those we are genuinely averse to and see ourselves as getting.

What is not plausible is the thought that a person who lacks many crucial goods deserves equally to be given access to these goods as they deserve to be told to focus on those of their genuine desires that they take to be satisfied. My first point, then, is that it is one thing to say that these things might make a person better off to just the same extent, and quite another to say that we have just as much reason to give them one of these things as we do to give them the other. Indeed, this person might deserve the right to be autonomous, and to have access to a basic education, but surely deserves these more than to be told to focus on those of their genuine desires that are satisfied. Thus we must be careful not to mix our judgments about what we have most reason to do for this person with our judgments about what would most improve this person's life.

Furthermore, at least part of our intuition here—perhaps even the part concerned with what the person deserves—may be tracking the fact that goods such as health, autonomy, education, and the like, allow those who have them to reach much higher levels of well-being than those who do not. Possession of these goods allows us not only to form and satisfy a greater range of genuine desires, but also to focus on these satisfied desires rather than on our many frustrated ones. The ability to read, for example, gives its possessors access to a great deal of information that may give rise to or help satisfy certain genuine desires. It may give rise to a genuine desire to read great novels, or help to satisfy a genuine desire to get to some location. Being healthy, in contrast, also allows us to focus on those of our genuine desires that we take to be satisfied, instead of on those that appear frustrated by a lack of health. The pain involved in an unhealthy state may make it much more difficult to enjoy other things. All of this suggests that giving a person access to these goods is more likely to improve the person's life, on the genuine desire theory of well-being, than simply telling them to focus on those of their genuine desires that they take to be

satisfied. The theory can then account for the intuition that we should be doing the former rather than the latter.

Finally, it may be worth mentioning that focusing on a particular one of our genuine desires that it is possible for us to see as satisfied, or even on any of our genuine desires that it is possible for us to see as satisfied, can be much harder than it sounds. If, for example, one is in great pain due to some disease, or a lack of food or water, it may be quite difficult to focus on the fact that one's genuine desire to be alive remains satisfied. Indeed, it might even be impossible, psychologically, to focus on this fact, if one's deprivation is great enough. Surely it will not be as easy as simply "choosing" to focus on it. Consequently, a subject who lacks basic goods such as health, autonomy, and education may have a harder time focusing on those of their genuine desires that they are able to see as satisfied. Giving a person access to these goods may then be a better bet for making the person better off, on the genuine desire theory of well-being, than simply telling them to focus on things that they both genuinely want and take themselves to have. So, again, the theory can account for our intuition that we should be doing the former rather than the latter.

5.2 Adaptive preferences

But now a related worry might seem to arise: suppose our oppressed subject not only formed a tendency to focus on those of her genuine desires that she takes to be satisfied, but also began to adapt her preferences or desires to her current conditions. More specifically, suppose she began to genuinely desire to have very little food, money, education, and very few rights or liberties. The genuine desire theory of well-being seems to imply of this person that she is doing very well, and indeed that her forming adaptive preferences is good for her. Intuitively, however, it seems clear that she is not doing well. After all, she is deeply

deprived, and the victim of grave injustice. She is genuinely desiring things that are surely bad for her. So how could it be said that she was doing very well, or that her adaptive preferences were good for her?

Well, for one thing, it is not entirely clear that the genuine desire theory of well-being implies that she is doing very well. That's because the objects of her enjoyment here, such as the inadequate amount of food, unclean water, or restrictive liberties available to her, do not have much objective worth. That is, there are very few (if any) reasons for anyone capable of genuinely desiring these things to genuinely desire them, especially considering the alternatives available to those in non-oppressive conditions. That being so, the enjoyment she derives from seeing herself as getting these things contributes much less to her well-being, on the genuine desire theory of well-being, than enjoyment derived from seeing herself as getting more objectively valuable things, and doing so more frequently, would. In that case, its seems plausible that she would not be doing very well on the theory, which therefore faces no worry here.

For another thing, admittedly, the genuine desire theory of well-being may sometimes imply that adaptive preferences might be good for one, or that they might make one better off. This depends on how we fill out the details of the case. Still, as I see it, any case where the theory has this implication will be one where the subject does indeed seem better off. To see this, consider first a case where a subject begins with a great life in conditions of oppression but no deprivation, and now has adapted preferences about her oppressive condition. She originally saw the oppressive conditions as undesirable, and she now sees them as desirable. Still, she experiences the same amount of enjoyment before and after this change, and some of the enjoyment she now experiences is derived from her adapted preferences. In this case, the genuine desire theory of well-being implies that, so

long as what she was enjoying before had at least some objective worth, this change in her preferences left her worse off. This is because she experiences the same amount of enjoyment, and at least some of the enjoyment she now experiences is taken in something with no objective worth. The theory thus secures what I take to be the correct result. Consider next a case where a subject begins with a bad life in conditions of oppression and deprivation, and now has adapted preferences about her oppressive condition. Again, she originally saw the oppression and deprivation as undesirable, but now sees them as desirable. In fact, she originally saw everything as undesirable, and so these are now the only things she sees as desirable. She now experiences at least some enjoyment in connection with these desires. In this case, the genuine desire theory of well-being implies that the subject is better off in virtue of her preferences adapting. This is because she now experiences at least some enjoyment, and while this enjoyment contributes to her well-being at a reduced rate, it nevertheless contributes something, making her at least somewhat better off. Again, I think, the theory secures the correct result.

These two cases are extremes in a wide range of possible cases of adaptive preferences. As they show, however, the genuine desire theory of well-being has the resources to plausibly explain many cases in this range. When someone experiences no enjoyment at all, and then experiences some enjoyment in virtue of their newly adapted preferences, they are at least somewhat better off. When someone experiences a great deal of enjoyment, and then replaces some of that enjoyment with enjoyment derived from their newly adapted preferences, they are likely to be at least somewhat worse off. In other cases, what will matter is the enjoyment before the preferences adapt, the enjoyment after the preferences adapt, the amount of enjoyment derived from the adapted preferences, and the objective worth of the things enjoyed before and after the preferences adapt. There is then

no clear answer to whether adaptive preferences make a given subject better or worse off. But I'm confident that the genuine desire theory of well-being will provide a plausible answer, given these further details.

5.3 What about the animals?

The genuine desire theory of well-being says, of any welfare subject, that all and only experiences in which a subject takes themselves to be getting what they occurrently genuinely desire are basically good for that subject, and all and only experiences in which they see themselves as getting what they are genuinely averse to are basically bad for that subject. It follows from this that, if a thing is not capable of genuine desire, or as seeing themselves as getting what they genuinely desire, that thing is not a welfare subject. That is, it follows that anything that is not capable of genuinely desiring, or of seeing these desires as satisfied, is not a thing to which our concept of well-being applies. Such things do not have any level of well-being; nothing can make their lives—if they have them—go better or worse for them. Thus if we have reason to think either that there is something that is incapable of either genuine desire or appearance of such desire's satisfaction, but that our concept of well-being seems to apply to, or else that there is something that is capable of both genuine desire and appearance of such desire's satisfaction, but that our concept of well-being does not seem to apply to, the genuine desire theory of well-being will face a significant worry. These things would serve as clear counterexamples to the theory.

If there are any such things, what might they be? The most obvious candidates seem to be animals. Surely we will want to say that non-human animals such as dogs, dolphins, cows, mice, and the like, are welfare subjects. After all, it seems natural to think there are things that are basically good and bad for them—that make their lives go better or worse for

them. These things might include eating when they are hungry, playing with other animals, getting exercise, and, on the other hand, being injured, suffering physical harm, and being confined. Other candidates for a counterexample might be mentally impaired or underdeveloped human beings. Both the mentally challenged and human babies certainly seem to be welfare subjects. They seem better off when they are happy, healthy, or having enjoyable experiences; and they seem worse off when they are unhappy, unhealthy, or suffering. An adequate theory of well-being should be able to account for these judgments.

Fortunately, the genuine desire theory of well-being can meet this demand, for all of these beings seem capable of both genuinely desiring and seeing themselves as getting what they genuinely desire. Dogs, dolphins, cows, mice, and other, more cognitively advanced non-human animals must be capable of having appearances—how else would they represent the world, or remember these representations? There is nothing abnormal about saying that, for example, the dog sees that you've thrown the ball, or the cow sees that there is grass. And these animals are nearly certain to be capable of genuine desire—they are moved to eat, drink, and have enjoyable experiences just as we are. The same seems true of human babies and many of the cognitively impaired: given they represent the world and are moved by their appetites, they appear capable of both having appearances and genuine desire. Each of these judgments I take to be quite uncontroversial. Indeed, I take them to be just as uncontroversial as the claim that these beings are all welfare subjects. We think of babies, many of the cognitively impaired, and many cognitively advanced non-human animals as being capable of well-being and ill-being. We also think of these beings as being capable of having appearances and genuine desires and aversions. We are correct in our first thought, I claim, in virtue of being correct in our second. The thoughts are equally uncontroversial because they amount to the same thing: to be a welfare subject is to be capable of having

genuine desires, genuine aversions, and appearances of these attitudes as satisfied. If that is right, the genuine desire theory of well-being can account for the intuition that these beings are all capable of welfare.

The more pressing problem for the theory, however, is beings that are less cognitively sophisticated. These beings might include humans who are incapable of having appearances or genuine desires, earthworms, or plants. Admittedly, it is difficult to imagine what life would be like as a human being who was incapable of having appearances and genuine desires, yet was also conscious. It is difficult to imagine what that sort of consciousness would be like. The same seems true of earthworms, and other, less cognitively sophisticated beings: if they are incapable of having appearances or genuine desires, it will be difficult for us even to imagine what their lives must be like. Nonetheless, when we consider what these lives must *lack*, it becomes clear, I think, that the lives could not be better or worse, in themselves, for the beings who live them. While we may lack the resources for imagining what life as (say) an earthworm might be like, we can perhaps go some way toward imagining this by thinking of what our own lives would be like if we lacked the capacity for genuine desire. If we never felt genuine attraction or aversion towards anything, what would our lives be like? As I see it, they would be devoid of goodness and badness; indeed, they would lack the capacity for well-being and ill-being.[58] After all, in this case, we would never enjoy or dislike anything. We would be affectively neutral, like robots. If, as I have argued, well-being requires enjoyment, it seems clear that our lives here would be incapable of well-

[58] This is the case, at least, in the sense of goodness that we use to pick out *the* good, or that which is in itself good. There is another sense of goodness, however, in which talk of what is good for beings like earthworms can be appropriate. This is the sense in which it might be good for (say) an earthworm to have food, or good for a knife to get sharpened. Goodness in this sense concerns the proper functioning of the thing in question. That we speak of the good of earthworms in this sense does nothing to suggest that earthworms are capable of happiness or welfare, however, just as our talk of the good of knives in this sense does nothing to suggest that knives are capable of happiness or welfare.

being. Similarly, if we never disliked anything, it is not clear how anything could be bad, in itself, for us. Our stance toward everything would be like our stance now toward the fact (supposing it is one) that there are an even number of stars in the universe, or toward some other trivial fact: utter indifference. It is difficult to imagine anything that could make lives of this sort go better or worse for those who lived them. This, as I see it, is because there are no such things. Though it may be tempting to think otherwise—to think that things such as health, knowledge, or the like, may be good for beings of this sort—this thought loses its allure the instant we recall that the beings who would possess these things would be thoroughly unimpressed by them. They would not be moved or care one lick about having them; if they lost them, they would think that that was just as well. What is more, the things would in no way position them to care about or be moved by anything else. In that case, the things don't seem to make their lives any better. And so, since there is nothing that could move them in this way, nothing seems good or bad for them.

Thus, if a being is incapable of having appearances or genuine desires or aversions, they are incapable of well-being and ill-being. While this may seem controversial, it really is not. It amounts to no more than the claim that if something is incapable of enjoying or disliking anything, nothing can be good or bad, in itself, for that thing. I do not see how this claim could plausibly be denied. It seems, then, that the genuine desire theory of well-being again secures the correct result: anything that is incapable of seeing itself as getting what it genuinely desires (not) to get is also incapable of welfare.

5.4 The Euthyphro objection

Simple (objective) desire satisfactionism tells us that the only thing that is basically good for us is the actual satisfaction of our intrinsic desires. This, it is sometimes claimed, gets the

relation between what is good for us and what we desire backwards. On this theory, something is good for us (if and only if and) simply because we intrinsically desire it and are in fact getting it. This suggests a picture on which our intrinsic desires in some way confer prudential value onto their objects, whatever those objects might be. Rather than things such as health, autonomy, and knowledge being basically good for us, then, this view implies that these things are good for us only in virtue of our intrinsically desiring them. Yet it seems clear that we could intrinsically desire some things that are no good for us at all. If, for example, a subject intrinsically desired that a stranger she had met was now experiencing a great deal of pain, and the stranger was, unknown to the subject, in fact experiencing such pain, it seems implausible that the subject's desire simply *makes* the occurrence of this state of affairs good for her. Indeed, if we suppose also that she would feel only shame and regret, or even that she would be affectively neutral, upon discovering that the stranger was in pain, it seems implausible that the occurrence of the stranger's pain is even good for her in the first place. The satisfaction of intrinsic desire, when it leads to unpleasantness, or is unknown, does not always seem to make a subject better off. This suggests that simple desire satisfactionism gets things the wrong way around: things are not good for us simply because we intrinsically desire them, but rather, we often intrinsically desire things because they are good for us. This is not unlike Socrates' claim that piety must be loved by the gods because it is holy, as opposed to holy because it is loved by the gods. The mere fact that we intrinsically desire something that has no worth does not somehow make the obtaining of that thing prudentially valuable. Yet, since we often intrinsically desire things that do seem to make us better off, it is easy to mistakenly conclude that it is in virtue of our wanting and getting them that they are good for us.

A similar objection may seem to apply to the genuine desire theory of well-being. After all, this theory claims that well-being consists in experience in which one takes oneself to be getting what one occurrently genuinely desires. And so, it may be thought, this theory too gets the relation between what is good for us and what we desire backwards. More specifically, it may be thought that the theory implies that occurrently genuinely desiring and seeing oneself as getting something confers prudential value on that thing. And this implication seems misguided: things are not good for us simply because we genuinely desire and see ourselves as getting them; rather, we genuinely desire them because they are good for us. And so, it seems, the genuine desire theory of well-being must be false.

This, however, is no objection to the genuine desire theory of well-being. Notice that the objection relies on the claim that our theory implies that occurrently genuinely desiring and seeing ourselves as getting something confers prudential value on that thing. Yet the genuine desire theory of well-being has no such implication. That's because the theory says that well-being consists in the *experience of* seeing oneself as getting what one occurrently genuinely desires. It does not say that well-being consists in the *objects* of those desires. Accordingly, the theory does not imply that occurrently genuinely desiring and seeing ourselves as getting something confers prudential value on that thing. Rather, it implies that occurrently genuinely desiring and seeing ourselves as getting something is a type of experience that is itself prudentially valuable. And, in fact, it implies that the objects of those attitudes partly determine the prudential worth of the experience—if they are more objectively valuable, the experience will have more prudential worth. Contrary to what this objection suggests, then, the genuine desire theory of well-being is able to accord with the fact that we sometimes genuinely desire things because we take them to be valuable; and that

this value can itself enhance the prudential worth of experiences in which we are focused on the fact that we genuinely desire and see ourselves as getting these things.

5.5 Paradoxical desires

Simple (objective) desire satisfactionism seems also to face the following sort of puzzle. Imagine a subject who has just one intrinsic desire: the desire to have a negative level of well-being. Since simple desire satisfactionism claims that well-being consists in the world being such that one is getting what one intrinsically desires, the theory seems to imply that, if this subject is getting what they intrinsically desire—if they have a negative level of welfare—then they are doing well, and if this subject is not getting what they intrinsically desire—if they do not have a negative level of welfare—then they are doing poorly. That is, it seems to imply that the subject has a negative level of welfare only if they have a positive level of welfare, and they have a positive level of welfare only if they have a negative level of welfare. That, it seems, is paradoxical. And so it seems a strike against the theory that it has this counterintuitive implication.

Can a similar objection be pressed against the genuine desire theory of well-being? Imagine a subject who has just one genuine desire: the desire to not have experiences in which they take themselves to be getting what they occurrently genuinely desire. The theory seems to imply that, if this subject takes themselves to be having such an experience, then they are doing poorly, and if this subject takes themselves not to be having such an experience, then they are doing well. However, so long as the subject takes themselves to be having an experience in which they see themselves as getting what they occurrently genuinely desire, the theory implies that they are also, to that extent, better off. And, so long as the subject takes themselves not to be having such an experience, the theory implies that they

are also, to the extent that they are averse to this, worse off. So, the theory in fact implies that, if the subject takes themselves to be having such an experience, they are thereby made to some extent worse off—since they occurrently genuinely desire not to have such an experience—and also made to some extent better off—since they nevertheless take themselves to be having such an experience. And the theory implies that, if the subject sees themselves as not having such an experience, then they are thereby made to some extent better off—since they occurrently genuinely desire to not have such an experience—and also made to some extent worse off—since they are nevertheless not having the relevant experience. Accordingly, the theory seems to imply that, when a subject's genuine desires are aimed at their own frustration, their perceived satisfaction will leave a subject, roughly speaking, no better or worse off. If one genuinely desires not to have prudentially valuable experiences, and one gets what one wants, this will entail that one has a prudentially valuable experience. It will also entail that one has a prudentially disvaluable experience. It seems plausible, then, that the result will be a wash. What's more, it's worth noting that the object of this desire—that they themselves do not, or for that matter that anyone does not, have prudentially valuable experiences—seems not to be one with much objective worth. This means that the increase and decrease in welfare that this affective oddball would get in virtue of having these experiences may in any case be quite small. So, the genuine desire theory of well-being rightly implies that, when a subject occurrently genuinely desires just that they have experiences in which they take themselves not to be getting what they occurrently genuinely desire, the result of any such experience on this subject's level of welfare will be somewhere between minuscule and nothing.

5.6 False beliefs and defective desires: an asymmetry?

Recall the two types of cases that disproved Feldman's claim that happiness precisely tracks welfare: cases of deception, and cases of immoral desire. These cases posed a problem for Feldman's view because, intuitively, possession of a false belief or immoral desire seems to negatively impact a subject's welfare, but such possession does not seem to impact the subject's happiness. It was with this intuition in mind that we expanded our theory of happiness into our theory of welfare. Here, we claimed, being well off to some degree requires more than being happy to that degree: specifically, it requires that what we genuinely desire (and see ourselves as getting) has some objective worth. This gives voice to our intuition that immoral desire seems to negatively impact a subject's welfare. But what of our intuition that false belief has this effect? Our theory of welfare does not seem to give voice to this intuition, despite its role in motivating us to reject Feldman's view. More specifically, it seems to imply that false belief, in itself, cannot make a subject worse off. In that case, it might seem that there is a certain kind of asymmetry between belief and desire, on our approach. After all, the genuine desire theory of well-being implies that the falsity of one's beliefs, in itself, cannot make one worse off, yet the immorality of one's desires, in itself, can make one worse off. In other words, it implies that the fact that one's beliefs are defective cannot negatively impact one's level of welfare, but the fact that one's desires are defective can. Given the way we reached the genuine desire theory of well-being, this might seem a problematic asymmetry.

But it is not. This alleged asymmetry is the result of a conflation of two senses of the word "defective." The sense in which a deceived subject's beliefs are defective is the familiar one in which they fail in their constitutive aim of truth. That is, the subject's beliefs are defective in the same way that any false belief is defective: they aim at the truth, and fail in

their aim. Yet a subject whose desires are immoral need not have defective attitudes in this sense. To have defective attitudes in this sense, the immoral subject must simply have desires that are not in fact satisfied—desires, that is, whose objects fail to obtain. After all, just as beliefs aim at truth, so desires seem to aim at satisfaction—satisfaction is their constitutive aim. So, for the immoral subject's desires to be defective in this sense would just be for them to fail to actually be satisfied. When we say that this person's desires are defective, however, we do not mean that they are not satisfied. Rather, we mean that they fail to meet some standard of morality or rationality. This person has not failed in their role as a desirer, but has instead failed in their role as a moral agent. It is in this sense that their desires are defective. Once this is recognized, it becomes clear that there is no problematic asymmetry between belief and desire on the current approach. On the contrary, there are two senses of "defective." And since the genuine desire theory of well-being is unconcerned with the actual satisfaction of one's genuine desires—what matters is just their perceived satisfaction—the theory implies that, just as the falsity of one's beliefs, in itself, cannot negatively impact one's level of welfare, so the actual frustration of one's desires, in itself, cannot negatively impact one's level of welfare. There is, accordingly, no problematic asymmetry here. In fact, there is a plausible symmetry: whether our beliefs are true and our desires are satisfied each depend on the way the world is; and, on the genuine desire theory of well-being, one's level of welfare does not depend on the way the world is—it depends only on how one takes the world to be. The theory therefore implies that the possession of beliefs and desires that fail in their constitutive aims, in itself, cannot negatively impact one's level of welfare. In light of what has been argued to this point, that seems the right conclusion.

5.7 Wrapping up

I conclude that the genuine desire theory of well-being is able to account in a satisfying way for each of the putative objections considered here. Not only does this theory offer an explanation as to why each of the three traditional types of well-being theory have proven so popular, but it also offers satisfying solutions to many of the problems faced by these accounts. No other theory of well-being on offer can accomplish this much. It seems, therefore, that the case in its favor is overwhelming: The genuine desire theory of well-being is to this point the most plausible such theory available.

6 Wrapping up and Looking Forward

That concludes my development and defense of the genuine desire approach to happiness and welfare. This process began by analyzing a recently popular approach to well-being, which claims that well-being consists in "enjoying the (objectively) good." We have seen that this approach is untenable, in large part because its attempt to restrict welfare to only enjoyment of a certain kind failed to mesh with the intuition that enjoyment of any kind has at least some goodness for the person who experiences it. This led us to examine several versions of hedonism, which were unified in their claim that all and only enjoyable experiences are basically good for us. One such view, offered by Fred Feldman, offered a template on which to build. This view claims that happiness, and indeed well-being, consist in all and only experiences of attitudinal pleasure. The most fundamental problem with this proposal is that attitudinal pleasure is not the most basic currency of our happiness or our well-being. As we have seen, the basic constituents of our happiness and welfare are all and only experiences in which we take ourselves to be getting what we genuinely desire. Feldman also suggests that the objects of our enjoyment do not make a difference to the extent to which an experience of enjoyment contributes to our well-being. As cases of deception and immorality show, however, what it is that we enjoy intuitively seems to make a difference to our welfare, and yet does not intuitively seem to make a difference to our happiness. This at least suggests that welfare cannot consist entirely in happiness, and indeed that what it is that we enjoy—or, what the precise objects of our enjoyment are—is the key factor separating these concepts. This led us to an account on which, while both happiness and welfare

consist in all and only experiences in which we take ourselves to be getting what we genuinely desire, the extent to which these experiences contribute to our welfare partly depends on the objects of our enjoyment, yet these objects are irrelevant to our happiness. Some of the things we enjoy provide us with more reasons for anyone (capable of desiring the things) to desire them than others, and enjoyment of a given intensity and duration will contribute a greater amount to our welfare to the extent that the objects of the enjoyment provide us with these reasons, yet will contribute the same amount to our happiness regardless of the objects of the enjoyment.

Taking a step back, we can see that the path we have taken to this point has been largely "top-down:" we have begun with broad ideas and concepts, such as pleasure and objective goods, and have analyzed them until we reached their more basic parts. This has led to the genuine desire approach to happiness and well-being. What I now want to do is adopt an alternative approach to our analysis of the concept of welfare, or adopt what we may call a "bottom-up" approach. My hope here is that, taking what we have learned from the top-down approach, we might be able to construct our theory of welfare by beginning with an analysis of its more basic parts. More specifically: by looking at our genuine desires, the role they play in our psychology, and what relationship they share to our evaluative intuitions, we might be able to construct a more illuminating version of what will ultimately amount to something very much like the genuine desire theory of well-being. This might even lead to insights into other important concepts, such as pleasure and "objective" goodness.

Of course, all of this remains, at this point, largely speculative. But, as I see it, the idea also seems quite promising. Indeed, I have a feeling that the analyses resulting from this approach will be at least as illuminating as those we have seen thus far.

In what follows, then, I will take up this "bottom-up" approach to analyzing well-being, and show how it can lead us to important insights into the nature of this and other concepts. In the following chapter, I will show how we can build a theory of well-being from genuine desire alone. This theory will be, in effect, "extensionally equivalent" to the genuine desire theory of well-being: it will reach all of the same conclusions about the well-being of particular subjects, though by a somewhat different route. Then, in chapter 8, I will show how this novel theory may help the hedonist avoid several objections that are often thought to undermine their view. Before getting to any of that, though, let me say a bit more about why there is any second part, and in particular any alternative analysis of the concept of well-being, included in this dissertation at all.

6.1 Motivating the alternative

While I have given some idea of what I aim to do in the second part of this dissertation, it will help for me to be a bit more explicit. I have mentioned that I want to adopt a "bottom-up" approach to analyzing the concept of well-being, where this stands in contrast to the "top-down" approach taken in part I. But what exactly does that mean? And, perhaps more importantly, what is the motivation for offering this contrasting analysis?

To answer both of these questions, it may be helpful for me to begin with a concession. Specifically: the genuine desire theory of well-being might appear to suffer from an issue that's not unlike one plaguing the very first sort of theory we analyzed at length, or the sort on which well-being consists in "enjoying the good." Those theories are motivated by the plausible thought that, while enjoyment is no doubt a key determinant to how well our lives go, there is an important sense in which we only seem to be doing well provided we are enjoying the "right" things. After all, a person who gets a great deal of enjoyment from

(say) watching innocent people being tortured does not seem to be doing as well as a person who gets enjoyment of an equal intensity and duration from acting virtuously. Intuitions of this sort motivate the claim that well-being consists in, not just any enjoyment, but only enjoyment that is taken in "objectively good" things. As we have seen, however, the support our intuitions lend to this claim is ultimately rather limited. It simply does not seem right to say that no amount of immoral or base enjoyment could contribute anything to the subject's well-being. The accounts imply, for instance, that an eternity spent enjoying morally repugnant things a great deal would be no better than no life at all. While some may be willing to bite this bullet, it seems to me at least as implausible as the implications of hedonism that the accounts were designed to avoid.

The genuine desire theory of well-being was also designed to account for certain of our intuitions about well-being, and may seem to account for them in a way that ultimately leads, in a similar way, to disappointment. The worry is somewhat subtler in this case, though it remains an attempt to force a theory to accord with our intuitions in a way that may seem problematic. To be specific: the worry is that we have effectively fixed or rigged our theory to generate all of the intuitively correct results regarding the well-being of particular subjects. We have done this by stipulating that any change in well-being that cannot be explained merely by the amount of enjoyment or unpleasantness the subject experiences—such as the changes involved in cases of deception and immoral enjoyment—can instead be explained by the objects of the enjoyment or unpleasantness. This may seem problematic because, if it is not ultimately the enjoyment or unpleasantness itself that explains the changes in well-being, the claim that the basic constituents of our well-being are all and only experiences of enjoyment may seem to lose some of its credibility. After all, if these experiences are the basic constituents of our well-being, they alone should presumably

be capable of explaining any intuitive change in well-being. Why should something that is not an essential ingredient of these experiences be capable of explaining some changes in well-being, if the experiences alone are the basic constituents of our well-being? The only answer seems to be: because if they do not explain the changes, nothing will. The appeal to the objective goodness of the things we genuinely desire may then begin to look more like a desperate attempt to rig the theory's implications, and less like a well-motivated argumentative move.

Now, I do not think that this objection seems nearly as forceful as the related objection leveled against theories claiming well-being to consist in enjoying the good. In fact, I am not entirely sure that I take this argument to present any worry for the genuine desire theory of well-being at all. After all, there might be an adequate response available. It might be responded, for instance, that contrary to the objector's suggestion, the objects of our enjoyment do not ultimately explain certain changes in well-being, but rather help to identify a third dimension, in addition to duration and intensity, along which an experience of enjoyment may vary. In that case, it might be that this third dimension of the experiences explains the relevant changes in well-being, and so there is no inessential ingredient of the experiences that sometimes explains the changes. In any case, it is not my aim here to defend this, or any, response to the current objection to the genuine desire theory of well-being. It is instead to simply note that this theory's appeal to the objects of our enjoyment in explaining how certain changes in well-being occur may appear to some to be a weak point of the theory. While my objection offered on behalf of these people has not left me—and, I suspect, will not leave others—entirely convinced of this weakness, I must admit that I am also not entirely convinced that there is no real worry here at all.

If there is indeed a problem in this vicinity, notice that the mistake leading to it will be similar to the one that led to the previously discussed problem for theories claiming well-being to consist in enjoying the good. More specifically, the mistake will lie in the attempt to force our theory to produce the intuitively correct results, only after recognizing that the theory cannot reach these results without some tweaking. We change our definition of the basic constituents of our well-being in response to certain of our intuitions, rather than try to explain the intuitions in a way that does not conflict with our extant definition of the basic constituents of our well-being. Though, as I have said, I am not convinced that this way of doing things is obviously problematic, I am also not convinced that it is unproblematic. I think it is then at least worth exploring the alternative route, which begins with our more basic definition of what well-being consists in, and hopes from there to find a unifying explanation for all of our intuitions about changes in well-being. Rather than rig our theory to avoid problems that we cannot find a more basic, unifying solution to, we might search for this more basic solution by taking a different route.

What then do I mean when I call this alternative approach "bottom-up?" I mean that we are beginning with more basic concepts, such as genuine desire, and constructing a broader theory from there. The hope is that looking at the role these desires play in our psychology may lead to a more illuminating picture than the one beginning from enjoyment and objective goods, and ending with attempts to fix whatever problems remain after reaching the more basic concepts. Building a theory of well-being via this "bottom-up" approach is the task of the following chapter.

Part II

7 Well-Being as Reliable Happiness

Consider a case. Imagine that you have an old friend, Ann, who you've learned never to rely on for anything important. If you ask her to bring something she will forget it, if you ask her to look after something she will neglect it. Time and again, Ann fails to help you when you call. And it is not that your requests are in any way unreasonable. Ann just can't be relied on.

Now suppose that you need to be rushed to the hospital, and the only viable way of getting there is by Ann taking you. Let's also stipulate that Ann will definitely come through for you this time: she is certain to get you to the hospital, and you know this. While we're waiting for Ann, I turn to you and (sincerely) ask: "can you count on Ann to come through for you this time?" Since we've stipulated that she will come through for you, there is a sense in which it would be inappropriate for you to say anything but Yes. But because Ann is generally so unreliable—conditioning you to resentfully expect her to let you down again—this answer, on its own, is likely to leave you with a bad taste in your mouth. As a general rule, Ann cannot be counted on. But because we've stipulated that she can be counted on in this case, the general rule would lead you astray. The current case is then one in which the general rule is not to be trusted. Still, when I ask you whether you can trust Ann to come through for you this time, you cannot help but feel the residue of the general rule's verdict: she cannot *really* be counted on, intuition suggests, but still you begrudgingly concede that she can be in this case.

In this paper, I want to argue that your situation in this case is analogous to a situation we face in the debate over the nature of well-being. To see what I mean by this, let me give a bit of background.

Recall that theories of well-being or welfare essentially aim to tell us which things are basically good and bad for us—or which things are the basic constituents of our well-being and ill-being. And according to

> *Hedonism*: all and only enjoyable experiences are basically good for a subject; all and only unpleasant experiences are basically bad for a subject; and an enjoyable (unpleasant) experience contributes to a subject's well-being (ill-being) just to the extent that the experience has a greater intensity or duration.[59]

Since hedonism claims that only experiences of a certain kind can directly contribute to (or detract from) a subject's well-being, the theory implies that nothing external to these experiences can directly influence a subject's well-being. The hope for the hedonist, then, will no doubt be that our intuitions accord with that implication, at least provided our intuitions are to be any guide to a theory's plausibility.

If that is right, then it is unfortunate for the hedonist that our intuitions as to whether things external to a subject's experiences can directly influence their well-being are at best mixed. The most pressing problem for the hedonist is cases of deception. Ever since Nozick's introduction of the experience machine, philosophers have taken the thought

[59] I will then continue in part II to follow Roger Crisp (2006, ch. 4) in formulating hedonism in terms of "enjoyment" rather than "pleasure," for reasons he notes. I will also continue to use "unpleasant" rather than "painful," for parallel reasons. On this picture, "pleasure" ("pain"), as we generally use the term, is just one kind of enjoyment (unpleasantness).

experiment to show that hedonism is false. Indeed, the view is often rejected in passing, alongside just a reference to the famous passage:

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading a book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? (Nozick 1974, p. 42)[60]

Most people, it has been said, would not, or should not, plug in. That tells against hedonism, since it reveals that we care about more than enjoyable experience, making hedonism, as a theory of what is basically good and bad for us, appear inadequate or shallow.[61]

This line of thinking has recently been challenged, however. A number of philosophers, some citing experimental data, have questioned whether our intuitions about the experience machine tell us anything about the truth or falsity of hedonism.[62] At the very least, this data seems to show that there is little agreement in our intuitions about the importance of the veracity of our experiences. But, even so, Eden Lin has recently argued

[60] See, for example, Griffin 1986 (pp. 9-10), Hausman 2012 (p. 79), Hurka 2011 (pp. 68-70), Kagan 2009 (p. 253), and Keller 2009 (p. 657). For twenty-five more examples—or what Dan Weijers calls "just a sample"—see Weijers 2014 (p. 530, fn. 1).

[61] Though the inference from "care about" to "has prudential value" has been challenged by Fred Feldman (2011).

[62] See, for example, De Brigard 2010, Weijers 2014, and Weijers and Schouten 2013. See also Hewitt 2010, Kawall 1999, and Silverstein 2000. A similar project of undermining our intuitions about the experience machine has recently been taken up by Ricard Kraut (2018), though he does so in the service of defending a view that he calls "experientialism."

that there is another sort of case involving the experience machine that we can all get behind:

> Consider two lives, A and B, that are experientially identical and thus identical with respect to the qualitative features, durations, and temporal distribution of the pleasures and pains they contain. The subject of A (call him Adam) spends his life in the real world, whereas the subject of B (call him Bill) is plugged into an experience machine for his entire life. A is a good life of the sort available to citizens of Western countries. Let us stipulate that at no point does Bill interact with, or receive any care from, other human beings: thus, the experience machine runs entirely on its own, without any human intervention. Indeed, at no point after Bill's birth is any person even aware of his existence. (Lin 2016, p. 321)[63]

Lin predicts that many of us will join him in having what he calls *the comparison intuition*, or the intuition that "A is at least somewhat higher in total welfare than B" (2016, p. 321). This judgment is also supported by the "fact that there appears to be something especially pitiful about Bill," and since "Bill warrants more pity than Adam, he must be lower in welfare" (Lin 2016, p. 321).[64]

I suspect that Lin is right that many of us will share his comparison intuition. And, if he is right, then that appears to spell trouble for the hedonist. That's because, as Lin notes,

---

[63] Lin considers a similar case in his 2014 (pp. 141-145), again in the service of discrediting hedonism (as well as monistic theories more generally). Lin's argument there depends on some assumptions about the nature of pleasure and desire that, as will become clear below, I believe are mistaken. See also Crisp 2006 (pp. 117-119).

[64] I will return to analyze this point about pity in chapter 9.

the veracity of the comparison intuition is inconsistent with the truth of hedonism. Since hedonism claims that well-being consists in enjoyable experience, it follows that two lives with identical experiences should have identical levels of welfare. And so, since A and B are made up of the same experiences, it seems the truth of hedonism is incompatible with the veracity of our intuition that A and B have distinct levels of welfare. Lin calls this "*the experientially identical lifetime comparison argument*—or *the comparison argument*, for short" (2016, p. 322). It claims

1. A is higher in welfare than B.
2. If hedonism is true, then A is not higher in welfare than B.
3. *Therefore*: Hedonism is false.[65]

Though simple, this argument has a good deal of intuitive force. Lin takes it to provide us with, not a knockdown argument against hedonism, but at least a pro tanto reason for rejecting the view.

This back and forth between the hedonist and her challenger can be seen as one instance of a more general puzzle in the debate over the nature of well-being.[66] We seem to have two sets of contrasting intuitions about the nature of welfare. On the one hand, it seems obvious that the quality of our subjective experience is a key determinant of how well our lives go. If we never enjoy anything, our lives cannot be good ones. And if we are in no

[65] Lin 2016 (p. 322).

[66] This puzzle is introduced in Kagan 1992 (pp. 180-187), and discussed in Kagan 1994 (pp. 313-317). See also Keller 2009 (pp. 657-663).

sense attracted to or pleased by things—if they in no way "resonate" with us—the things seem to do little to improve our lives.[67] If, for example, I hope that you do well in life, but I then never think of you again, your going on to do well hardly seems to impact how well my life goes.[68] Yet if I come to find that you have gone on to do well, and I am pleased by that fact, your going on to do well suddenly seems relevant to my well-being. So if something is to impact my well-being, it seems it must impact *me*, in the sense of entering my consciousness and impacting my mental state. Such intuitions tell in favor of "mental state" or "experiential" theories, like hedonism. But on the other hand, it seems that things "external" to our subjective experience can also affect our well-being. Lin's case, for example, seems to show that whether we are in contact with reality matters to our well-being. Here we might also point to goods like health, autonomy, and knowledge: surely these things are good for us, regardless of how we feel about them. More generally, the quality of our "objective" life circumstances (or of relevant "states of the world") also appear to be a key determinant of how well our lives go. This leads us away from views that focus entirely on the "subjective" or "internal" features of our lives, like hedonism, perhaps toward "hybrid" views that combine subjective and objective elements, or even toward a full-fledged objective list account.[69]

In this chapter, I want to argue that the latter set of intuitions here are leading us astray. By defending hedonism against some recent attacks, I will show that we do not need

[67] As we've seen, this thought is sometimes called "internalism about well-being," or "the resonance constraint" (ch. 1.2).

[68] An example made famous by Parfit (1984, p. 494).

[69] Indeed, as we have seen, our conflicting intuitions here seem to be the main source of motivation for hybrid views on which well-being consists in "enjoying the good" (ch. 2). They also seem to motivate a recent objective list account. See Fletcher 2013.

to look at anything other than the subject's enjoyable experiences to determine their level of well-being. In the process, I will wed hedonism to a particular metaethical theory of welfare, leading to a picture of the nature of well-being that plausibly fills very many of hedonism's perceived gaps. On this picture, any "objective" or "external" life-circumstance that appears to affect how well a subject's life goes is in fact just a reliable indicator of future enjoyable or unenjoyable experience.[70] More precisely, it is just a feature of the subject's life that reliably indicates to us—the particular evaluators—that the subject has some enjoyable or unenjoyable experiences in their future. Since (e.g.) lack of contact with reality has proven to us to be a reliable indicator of future unenjoyable experience, the idea of life on the experience machine leaves us with a bad taste in our mouths. Much like in the case of Ann, however, our stipulation that the general rule will lead us astray ensures that this bad taste is of little importance in assessing cases like Lin's. Just as Ann has proven herself generally reliable with respect to letting you down, so enjoying yourself while failing to track reality has proven itself generally reliable with respect to future pain. When our general rules are stipulated to fail, the bad taste this leaves should be ignored. All of this leaves us with a picture on which well-being is a kind of temporally robust enjoyable experience—or a kind of *reliable happiness*.[71]

The chapter will proceed as follows. In the first section, I will argue that Lin's comparison argument does not give us a pro tanto reason to reject hedonism, and that the

[70] My talk of features of lives being "reliable indicators" of future enjoyable experience is inspired by Edward Craig's (1990) talk of "indicator properties" of potential informants in his analysis of the concept of knowledge. A property of this sort is meant to be "such that if the informant possesses it he is (at least) very likely to have a true belief on the matter" (Craig 1990, pp. 18-19). In my case, the relevant feature of the life is meant to be such that if the subject possesses it she is (at least) very likely to have some enjoyable or painful times ahead of her. Similar projects in epistemology that have influenced my thinking here include Dogramaci 2012 and Reynolds 2017.

[71] Here I assume that happiness consists in enjoyable experience, which is a point I argued for in chapter 3.

comparison intuition can, contrary to Lin's claims, be unproblematically debunked by the hedonist. In the second section, I will consider an argument that advances on Lin's claims, and suggests that we adopt a "bipartite" characterization of well-being. This way of characterizing well-being apparently allows us to see that our two sets of conflicting intuitions are each telling us something important about the nature of well-being. I explain where this proposal goes wrong, and thereby show that the hedonist's response to Lin can be generalized to account for any feature of a life, including any "external" or "objective" feature, that appears to affect the life's level of welfare. My aim, then, is to show that the limits of well-being are no more than the limits of enjoyable experience. I will conclude by discussing the connection between the view argued for here and the genuine desire theory of well-being.

7.1 A hedonistic metaethic

After introducing his comparison argument, Lin offers a challenge to it on behalf of the hedonist. Lin suggests that the hedonist might object that the comparison intuition can be undermined by way of an evolutionary debunking argument. If the hedonist can show that our comparison intuition has been formed in a way that strips it of its legitimacy, and this explanation also tells in favor of hedonism, then the hedonist may be able to evade Lin's charge altogether. That is, if hedonism predicts that we would have this intuition, and can also explain why the intuition is not to be trusted, it seems that the hedonist will have all they need to refute Lin's argument. So long as the comparison intuition is not to be trusted, it's clear that its propositional content should not be used as a premise in an argument. And without that content, Lin's comparison argument plainly does not go through.

Debunking arguments of this sort have been attempted by a few hedonists, all of whom argue along similar lines.[72] The most detailed such proposal, though, comes from Matthew Silverstein. Silverstein appeals to some claims made by Brandt and Railton, who tell a plausible story of how we come to acquire desires.[73] The idea is that when a desire leads to enjoyable experience, the desire is positively reinforced, making us more likely to keep hold of it and to act on it in the future. In contrast, when a desire leads to unenjoyable experience, the desire is negatively reinforced, making us less likely to keep hold of it and to act on it in the future. In this way, we have been conditioned to desire in accordance with enjoyable experience: we tend to desire things that have led us to enjoyable experiences, and to be averse to things that have led us to unenjoyable experiences. Importantly, as the so-called paradox of hedonism shows us, those who aim directly at enjoyable experience are often less likely to achieve it than those who aim at things that themselves tend to lead to enjoyable experience. The enjoyable experiences associated with friendship, for example, are not attainable by someone who intrinsically desires only those enjoyable experiences, because any relationship formed purely on the basis of that desire will not constitute a friendship. It then stands to reason that the conditionalization process would leave us with intrinsic desires for things that were not themselves, but that themselves tended to lead to, enjoyable experiences. All of our intrinsic desires would then be explainable in terms of their proven tendency to lead to enjoyable experience.

[72] See Crisp 2006 (pp. 120-122), Hewitt 2010 (pp. 343-347), and Silverstein 2000. Lin notes that his objections apply equally to all three of these arguments.

[73] Brandt 1979 (p. 100) and Railton 1989 (p. 167). For empirically-based defenses of the claim (or something very close to it), see Arpaly and Schroeder 2014 (ch. 6) and Schroeder 2004 (chs. 2, 3, and 4).

The next step in Silverstein's argument is to claim that our evaluative intuitions are informed by our intrinsic desires. If that is right, then all of our evaluative intuitions about things are explainable in terms of our having been conditioned to see those things as leading to enjoyable or unenjoyable experiences. That is, since our evaluative intuitions are a product of the proposed conditionalization process, they can be explained in terms of it.

Now here is how Silverstein takes this point to relate to the experience machine. One thing that we appear to intrinsically desire is that our experiences are veridical, or that they track reality. This would explain our having the comparison intuition, since our intuitions are just reflections of our intrinsic desires, and lives on the experience machine are not getting something that we intrinsically desire. Still, Silverstein thinks, we only have this intrinsic desire to track reality—or this aversion to being deceived—due to its having proven itself to be conducive to enjoyable experience. After all, in the real world, being deceived almost always leads to experiencing pain: think of someone who is cheated on by their partner, or secretly ridiculed by their friends. Importantly, though, in cases like Lin's, failure to track reality is *not* conducive to unenjoyable experience. Indeed, it does not matter whether one tracks reality, because one will have the same experiences either way. So, while we have been conditioned to think that, in Lin's case, B will include less enjoyable experience than A, we know (by stipulation) that this thought is wrong: they will include all the same experiences. The conditionalization process has then left us with a general rule that, in cases like this one, leads us astray. Thus when combined with this plausible story of the conditionalization of intrinsic desire, and the claim that our evaluative intuitions reflect our intrinsic desires, hedonism seems to predict that we would have the comparison intuition, and that this intuition would be misguided. The hedonist then seems to have a plausible way of undermining Lin's comparison intuition, and thus his comparison argument.

Lin, however, takes Silverstein to have shown too much: the argument undermines, not just the comparison intuition, but all of our evaluative intuitions. After all, Lin notes, Silverstein's argument runs roughly as follows: our evaluative intuitions reflect our intrinsic desires; our intrinsic desires (such as our intrinsic desire to track reality) are formed in such a way that we have no reason to expect their objects to be basically good for us; therefore, our anti-hedonistic intuitions should be distrusted.[74] Yet these two premises seem to yield a stronger claim: that *all* of our evaluative intuitions should be distrusted. As Lin notes, the fact that our intrinsic desires happen to have been conditioned to promote enjoyable experience does not make our hedonistic intuitions any more trustworthy than our anti-hedonistic ones. For our specifically hedonistic intuitions to be trustworthy, our possession of them must have something to do with the (putative) fact that enjoyable experience is the sole basic good. And since, on Silverstein's picture, the method by which we acquire our intrinsic desires shares no connection to the truth of hedonism—we just happen to be conditioned in this way—we have no reason to think that the method is leading to reliable evaluative judgments. Lin concludes: "The story about how pleasure explains our desires plays no role besides supporting the claim that, given how our intrinsic desires are formed, there is no reason to expect their objects to be basically good for us" (2016, p. 329). Thus, without a connection between this conditionalization story and the truth of hedonism, Silverstein's argument gives us no reason to think that our hedonistic intuitions remain unscathed. And this means that all of our intuitions are undermined.

What the hedonist needs, then, is an explanatory connection between the story of the conditionalization of our intrinsic desires and the truth of hedonism. Silverstein attempts

[74] Lin 2016 (p. 329).

to establish such a connection, appealing to the fact that only hedonism "explains what would otherwise appear to be a mysterious coincidence," namely the fact that all of our intrinsic desires "point towards" enjoyable experience (2000, p. 297). Such desires point toward enjoyable experience, Silverstein thinks, because enjoyable experience is the sole basic good. Pointing to Silverstein's claims here, Lin writes:

> Silverstein does not tell us *how* the truth of hedonism could explain the truth of the psychological story about desire-formation that he proposes. … I don't see how this particular explanation would go. Unless we posit a benevolent deity, wouldn't it be a mysterious coincidence that all of our desires are reinforced in accordance with how well they lead to the sole basic good? If we believed that our desires are reinforced in accordance with whether they lead to the propagation of our genes, we wouldn't conclude that the propagation of our genes is the sole basic good — and rightly so. Thus, the impression that Silverstein's story supports hedonism is an illusion. (2016, p. 330)

In the absence of an account of how the truth of hedonism explains the truth of the conditionalization story, we have no reason to think that the proposed conditionalization process is leading to reliable evaluative judgments. We have just two (putative) facts: that enjoyable experience is the sole basic good; and that our intrinsic desires are conditioned to "point toward" enjoyable experience. Without an explanatory connection between these

facts, their truth remains a mysterious coincidence—and a convenient one, for the hedonist.[75]

As I see it, however, the matter here is not all that mysterious. All the hedonist must do is posit a constitutive connection between enjoyable experience and intrinsic desire.[76] After all, if one were partially constituted by the other, then there would be little mystery as to how the proposed relation between the two came to be. What's more, we have already seen that the most plausible version of hedonism is one on which intrinsic desire is constitutive of enjoyable experience. We have seen, more precisely, that sometimes when we intrinsically desire things, we are genuinely attracted to the things—we are not merely disposed to get them, but are enthused or excited by them. This is the sort of desire we have called *genuine* desire (chapter 3.2). On the picture I have suggested, enjoyable experience *just is*

[75] As Lin notes, his challenge here parallels one made by Sharon Street (2006) to value realists. In the latter case, the two (putative) facts are: that there are evaluative truths that hold independently of our attitudes towards them (or, that value realism is true); and that our evaluative beliefs have been conditioned (via the forces of natural selection) to "point toward" those truths. Street's point, like Lin's, is that without an explanatory connection between the facts, their truth seems quite improbable—and quite convenient, for the value realist.

[76] Though I see this suggestion as plausible quite apart from its usefulness in responding to Lin's argument, I owe my recognition of its use here to Kieran Setiya's (2012, chs. 2-3) rigorous discussion of the connection between ethical fact and belief. Setiya argues that ethical knowledge (or the non-accidental reliability of our beliefs in ethics) is possible only if our moral beliefs are constitutively bound to the moral facts. Likewise, I think that the non-accidental reliability of our desires in according with enjoyment (or the good)—and thus the non-accidental reliability of our evaluative judgments in according with hedonism—is possible only if our desires are constitutively bound to enjoyment (or the good).

experience in which we take ourselves to be getting what we genuinely desire.[77,78] That is, it is just experience in which we take something that we are enthusiastic or excited about to obtain. This suggestion is supported by the fact that any experience we count as enjoyable seems to be experience in which we take ourselves to be getting something that we are genuinely attracted to. Activities as diverse as eating, playing, reading, relaxing, and working are all candidates for being enjoyable, and what seems to make them enjoyable is our genuine attraction to doing them. Similarly, when we take pleasure "in" things, such as the fact that our favorite team has won, it seems plausible that we are just taking something whose obtaining we genuinely desire, to in fact obtain.[79] More generally: to test this view, consider any enjoyable experience that you've ever had. Was it also an experience in which you took yourself to be getting something that you were genuinely attracted to? I'm willing to bet that it was.

[77] Admittedly, it is a bit awkward to say of something that we already see ourselves as getting that we genuinely desire to get it. That's because desire is generally seen as being in some important sense an essentially prospective attitude: it aims to fill a gap that we see in either our current or (potential) future circumstances. Thus if we see no such gap in our current circumstances—if we take ourselves to be getting the thing we genuinely desire—it will seem odd to say that we continue to genuinely desire to get the thing at that precise time, because we will see no gap that must be filled at that time. As I see it, though, this problem is more terminological than theoretical. The more natural term to describe our attitude in these cases, I think, is "like:" when our genuine desire leads to enjoyment, we at that point come to like the thing rather than yearn for it. This is no problem for the view I'm proposing, though, because I also think that to like something just is to both genuinely desire and currently see ourselves as getting the thing. Since liking is defined in terms of genuine desire, the latter is the more basic attitude, and so there is no worry of the current account being incapable of plausibly explaining (or, explaining in terms of genuine desire) why it sometimes seems odd to talk of enjoyment in terms of desire.

[78] As I've mentioned, Heathwood (2006, 2007) suggests a similar picture, though he defines sensory and "attitudinal" pleasure in terms of intrinsic desire. Given Heathwood's (2019) more recent arguments, though, it seems plausible that he too accepts something like the picture I'm proposing.

[79] Notice that pleasure of this sort, which I take to be what Feldman has in mind when discussing attitudinal pleasure, requires a belief: one cannot take pleasure in the fact that p without believing that p. In contrast, to experience just any enjoyment, on the current picture, requires only a "seeing-as:" one needs only to see oneself as getting what one genuinely desires to experience enjoyment.

Suppose then that enjoyable experience just is experience in which you take yourself to be getting something that you're genuinely attracted to.[80] How does this help the hedonist with respect to Lin's argument? Well, if genuine desire is constitutive of enjoyable experience, and enjoyable experience is the sole basic good, then there will be nothing coincidental about the fact that our genuine desires are conditioned to point toward enjoyable experience, or toward the good. When our genuine desires lead to enjoyable experience, they are positively reinforced, and when they lead to unenjoyable experience, they are negatively reinforced. We therefore tend to genuinely desire things, and see them as good, in accordance with how well they have led us to enjoyable experience, or to the good.

Here is another way of looking at it. When we genuinely desire something, we have a kind of affective investment in the thing. We see the good in the thing, and are thus enthused or excited by the thing. Experience in which we take ourselves to be getting the thing is the return on that affective investment. We not only see the good in the thing, but realize that good by taking ourselves to be getting the thing. Now if we receive the return on our investment, we will be more likely to make similar investments in the future. And if our investment does not pan out, we will be less likely to make similar investments in the future. In this way, we are conditioned to affectively invest more in things that tend to lead to returns on our investments, and invest less in things that tend not to lead to such returns. Our genuine desires are thus reinforced in accordance with how well they lead to enjoyable experience. And since they aim, at least indirectly, at enjoyable experience, there is nothing

[80] Of course, this assumption is not uncontroversial. Theories of pleasure are commonly divided into two types (see, for example, Sumner 1996 (pp. 87-92) and Lin 2018). One set of theories, which are usually called "internalist" or "phenomenological" theories, define pleasure in terms of a positive feeling tone. The other set of theories, which are usually called "externalist" or "attitudinal" theories, define pleasure in terms of a special sort of pro-attitude on the part of the subject. The approach I'm advocating is thus an externalist or attitudinal one. I have defended it in chapters 3 and 4.

mysterious about this connection. Our genuine desires aim at the good, and are reinforced in accordance with how well they attain it.

So it need not be, as Lin suggests, a mysterious coincidence that our genuine desires are reinforced in accordance with how well they lead to the sole basic good. Nor must the hedonist claim, as Lin also suggests, that it is in virtue of the fact that our genuine desires are reinforced in accordance with how well they lead to enjoyable experience that enjoyable experience is the sole basic good. He notes: "If we believed that our desires are reinforced in accordance with whether they lead to the propagation of our genes, we wouldn't conclude that the propagation of our genes is the sole basic good — and rightly so." On that much we can agree: how our genuine desires are reinforced—at least, considered in isolation—does not tell for or against hedonism. Instead, the truth of hedonism seems to tell in favor of the proposed story of how our genuine desires are reinforced. If the current picture of hedonism is correct, then our genuine desires are constitutively bound to enjoyable experience, or to our good. In that case, it is no wonder that our genuine desires are reinforced in accordance with how well they lead to enjoyable experience. Just as we are conditioned to believe in accordance with the truth, so we are conditioned to genuinely desire in accordance with the good.

Together with the claim that our evaluative intuitions reflect our genuine desires, this version of hedonism and the plausible story of how our genuine desires are conditioned yield the conclusion that Lin's comparison intuition is not to be trusted. On this picture, our genuine desires have been reinforced in accordance with how reliably they have led to enjoyment. Our evaluative intuitions, as reflections of these desires, are then tracking which things have reliably led to enjoyment, which in turn are just the things that, in virtue of their reliably leading to enjoyment, we see as reliably leading to the good. In this way, our genuine

desires each constitute a kind of general rule, according to which the object of the desire is a reliable indicator of future enjoyment. While we have been conditioned to accept as one such general rule that tracking reality will reliably lead us to more enjoyable experience—and so that A will contain more enjoyable experience than B—in this case the rule leads us astray. Since it is stipulated that A will not experience more enjoyment than B, the case is one in which our general rule is guaranteed to fail. Accordingly, our general rule in this case should be ignored, and so too should our intuition resulting from the rule's perceived violation. This debunks the comparison intuition, and does so without threatening our other evaluative intuitions. That's because we still have good reasons to think that most all of these intuitions are leading us reliably to a greater amount of enjoyable experience, given that the genuine desires that they reflect have been reinforced in accordance with how well they have led us to such experience in the past. Put differently: our genuine desires reflect their reinforcement, and their reinforcement confirms their reliability. When we know that their reliability has been undermined—as in Lin's case—we are justified in distrusting the corresponding intuitions. But when we have no reason to doubt their reliability—as in most any other case—we have no reason to distrust the corresponding intuitions. Contrary to Lin's suggestion, then, we do have good reasons to think that the things we genuinely desire are (at least derivatively) good for us. And so debunking one of our intuitions in this regard need not lead us to debunk them all.

It seems, therefore, that the hedonist has all of the resources needed to undermine Lin's comparison argument. To do so, they must simply claim that enjoyable experience just is experience in which we take ourselves to be getting what we genuinely desire, that our genuine desires are reinforced in accordance with how well they lead to enjoyable experience, and that our evaluative intuitions reflect our genuine desires. The truth of

hedonism would then predict that we would have the comparison intuition, and imply that the intuition is ultimately not to be trusted. And it would do this without casting doubt upon all of our other evaluative intuitions. Lin's comparison argument therefore presents no worry to the hedonist.

From this the following picture emerges. Enjoyable experience is the good, and we genuinely desire things in accordance with how well they lead us to more enjoyable experience. Since our evaluative intuitions are reflections of our genuine desires, and these genuine desires have been conditioned in accordance with how well they have led us to more enjoyable experience, it seems that our evaluative intuitions will reflect which things have reliably led us to more enjoyable experience, and the extent to which they have. If all that is right, it suggests the following test: if a subject intuitively seems to some degree better or worse off in virtue of having some feature, then that feature is to that degree a reliable indicator, to the evaluator, of future enjoyable or unenjoyable experience for the subject—that is, the evaluator's genuine desires have to that degree been conditioned in accordance with that feature's leading to enjoyable or unpleasant experience. In the following section, I will argue that this test gives the hedonist a plausible way of accounting for all of our seemingly anti-hedonistic intuitions. Whenever there is a feature of a life, such as an "external" or "objective" life-circumstance, that affects our intuition about a subject's welfare, that feature has proven itself to us to be a reliable indicator of future enjoyable or unenjoyable experience for the subject.

7.2 Hedonism and the limits of well-being

The comparison argument concludes, on the basis of the content of our comparison intuition and the fact that this content is incompatible with hedonism, that hedonism is false.

Yet Chad Stevenson takes our comparison intuition—and other anti-hedonistic intuitions like it—to support a more radical conclusion. This is the conclusion that we are often "considering multiple inputs when making welfare judgments," and that there are in fact two different "dimensions" of well-being. One dimension captures our "subjectivist" and hedonistic intuitions, while the other captures our "objectivist" and anti-hedonistic intuitions.[81] This way of characterizing well-being allows us to see what is correct about each of these sets of intuitions, without denying the legitimacy of either.

To reach this bipartite characterization, Stevenson appeals to the following pair of cases.

> Peter is happy, upbeat and generally in a good frame of mind. Yet events in Peter's life are poor, e.g. family death, relationship breakup, demotion at work.

> Paul has a loving family, job security and lives comfortably. Yet Paul lacks fulfillment. Paul is not happy or satisfied.[82]

It is, Stevenson thinks, at least not obvious which of the two subjects, Peter or Paul, is better off. Indeed, he adds: "it seems right to say that both are better off, and worse off, than the other, just in different ways" (Stevenson 2018, p. 392). More specifically, he thinks that it seems right to say that while Peter's life is not *going* well, he is still *doing* well; and that while Paul's life is *going* well, he is not *doing* well. The idea, then, is that a life might be marked by

[81] These "dimensions" also roughly track the distinction that Kagan (1994) notes between *me* and *my life*.

[82] Stevenson 2018 (p. 392).

good subjective experiences, yet the external circumstances (or the state-of-the-world) might not be favorable for the subject of the life; and a person's external circumstances might be favorable for them, yet their subjective experiences might be largely unpleasant.

Generalizing on this idea, Stevenson offers a bipartite characterization of well-being:

> *Doing-well*: an individual is doing well if and only if, then only in so far as, their subjective experience and attitudes towards their lives are of a particular set or kind.
>
> *Going-well*: an individual is going well if and only if, then only in so far as, a particular set or kind of life-circumstances obtain for them.[83]

Stevenson applies this characterization to Lin's case, writing: "while both are doing just as well, A's life is going better than B's. So A's welfare increases along both dimensions of welfare, whereas B's is only along one" (2018, p. 392). The characterization thus provides us with a "linguistic framework" to make welfare judgments, like this expanded version of the comparison intuition, "along two dimensions of assessment" (Stevenson 2018, p. 392).

Stevenson also takes this characterization to help us explain why our intuitions about cases involving the experience machine often conflict. The suggestion here is that, since we are considering "multiple inputs" when making welfare judgments—doing-well and going-well—we sometimes weigh these inputs differently, and thus disagree about "where the greater welfare pay-off lies" (Stevenson 2018, p. 395). In the cases of Peter and Paul, for example, which life seems better depends on which dimension of well-being we take to be

[83] Stevenson 2018 (p. 392).

weightier. Similarly, cases involving the experience machine often involve a choice between a life (on the experience machine) that would be higher along the doing-well dimension though lower along the going-well dimension, and a life (in the real world) that would be lower along the doing-well dimension though higher along the going-well dimension. On this picture, disagreements over which life is better are simply disagreements over how we should weigh tradeoffs between the dimensions. By recognizing this bipartite characterization, then, we can make sense of a major puzzle concerning our intuitive welfare judgments.

As may be expected, I think that this characterization of well-being is largely misleading. It takes as evidence for a separate "dimension" of well-being many of the intuitions that I believe are erroneously leading us away from hedonism; and it lists the only other such intuitions alongside enjoyable experience. This includes intuitions concerning "objective" or "external" life-circumstances, as well as "subjective" features (besides enjoyment), that appear to affect how well our lives go. Showing where this characterization of well-being goes wrong will therefore be instructive. It will teach us how the hedonist should account for any anti-hedonistic intuition.

As I see it, there is something important that Stevenson gets right. This is the idea that our intuitions reflect our calculations of prudential payoffs and tradeoffs. When making comparisons between lives on and off the experience machine, I believe that we are calculating the amount of future enjoyable experience that each life appears set to include. Judgments about which life contains more welfare are then judgments about which life has more properties that we have been conditioned to genuinely desire, and to thus see as leading to the good. Notice that Stevenson's examples lend credence to this suggestion. Recall that Peter, whose life was not supposed to be "going well," has endured a family

death, relationship breakup, and demotion at work. What all of these things have in common is that they have proven themselves to us to be extremely reliable indicators of future painful experience. Indeed, they are all things that we tend to be genuinely averse to, and that have proven themselves not to admit of any kind of quick fix. Thus the thought that things are not "going well" for Peter appears to be little more than the thought that Peter seems set to have some painful times ahead of him. Recall next that Paul, whose life was supposed to be "going well," has a loving family, job security, and lives comfortably. Again, what all of these things have in common is that they have proven themselves to us to be extremely reliable indicators of some kind of future experience, though in this case they are indicators of future enjoyable experience. They are all things that people are nearly universally genuinely attracted to, and that have proven themselves to reliably admit of a great deal of enjoyable experience in the future. Hence the thought that things are "going well" for Paul seems to be little more than the thought that Paul seems likely to have some enjoyable times ahead of him.

Similarly, Stevenson suggests that a subject may be "doing well" in virtue of her attitudes toward her life being of a particular kind. Yet this condition, too, simply picks out certain features of a life that have proven themselves to us to be reliable indicators of future enjoyment. That is, the features of a life besides enjoyment that are relevant to the doing-well "dimension" of well-being share with the "objective" or "external" features of a life that they are simply reliable indicators, to us, of future enjoyable experience. Again, Stevenson's examples tell in favor of this suggestion. Recall that Peter is happy, upbeat, and generally in a good frame of mind. Though these phrases are obviously somewhat ambiguous, it seems clear enough to me what they're getting at. They are suggesting that the subject not only regularly experiences enjoyment now, but also has a kind of internal constitution whose features have proven themselves to us to reliably indicate enjoyment for the subject in the

future. Peter appears to us, in other words, to be disposed to have enjoyable experiences in the future, in virtue of his being happy, upbeat, and in a good frame of mind. So while these things are features of Peter's subjective experience, still they are things that we are nearly universally genuinely attracted to, and that have proven themselves to us to reliably indicate future enjoyment. Recall next that Paul lacks fulfillment, and is not happy or satisfied. Again, these things all have in common that they suggest to us either that Paul is going through some painful times now, or that he seems likely to have some painful times ahead of him. Fulfillment and satisfaction (at least with one's life) are generally not feelings that return quickly after they are lost. Losing them suggests a deeper problem, which suggests painful times are ahead. Thus while these are all features of Paul's subjective experience, they are still things that we are genuinely averse to, and that reliably indicate to us that Paul will have some painful experiences in the future.

The point generalizes: any feature of a life besides enjoyment that appears to affect how well a life goes can be explained by the hedonist in terms of its having proven itself to us to be a reliable indicator of future enjoyable or unenjoyable experience. All of the relevant features of a life that seem to positively impact it are things that we see as leading to the good: they are all things that we are genuinely attracted to, and that have proven themselves to reliably lead to enjoyment. And all of the relevant features of a life that seem to negatively impact it are things that we see as leading to the bad: they are all things that we are genuinely averse to, and that have proven themselves to reliably lead to unpleasantness. Since our evaluative intuitions are, on this picture, reflections of our genuine desires and aversions, and these desires and aversions have been reinforced in accordance with how well they have led to enjoyment, our evaluative intuitions also reflect that reinforcement, and are in this sense predictions of future enjoyment. The intuitions are, in other words, tracking the features of

lives that have reliably led us to enjoyment, and the extent to which they have. We need not, then, posit any basic prudential goods other than enjoyable experience to make sense of our many intuitions concerning the nature of well-being. And we need not take the limits of well-being to be any more expansive than the limits of enjoyable experience.

We are now in position to identify what I take to be one of the most crucial and common mistakes made throughout the debate over the nature of well-being. This mistake is deriving a basic good from a future basic good. That is, it is taking as basically good for us things that have merely proven themselves to us to be reliable indicators of future enjoyable experience. To claim that there are "objective" features of our lives "external" to our experience, or certain "subjective" stances towards our lives, that can be basically good for us, is to make this mistake. This impugns arguments appealing to veridical experience, posthumous goods, knowledge, achievement, autonomy, "objective" desire satisfaction, and many more besides. These things all seem good for us, but this is only because they've proven to reliably lead to more enjoyable experience. The current approach then explains why we have for so long approximated the truth by appealing to these things. Theory after theory appeals to things that tend to lead to enjoyable experience. The best explanation for this, I think, is the current one. Well-being has its source in enjoyable experience, and our welfare judgments reflect which things have proven themselves to us to reliably lead to enjoyable experience.

This leaves us with a more nuanced picture of the nature of well-being. On the one hand, there is a sense in which it tells us that the hedonist was right all along: all and only enjoyable experiences can directly contribute to our well-being. But on the other hand, it tells us that our welfare judgments track which things have proven themselves to us to be reliable indicators of future enjoyable experience. So it would be somewhat misleading to say

that our well-being is really just a matter of enjoyable experience. It is also, in a sense, a matter of future enjoyable experience. Still, we should be careful to note that future enjoyable experience, as such, does not contribute to our well-being. Our well-being is constituted by enjoyable experience—experience that we actually have. Our welfare judgments aim to track future enjoyable experience—experience that we seem set to have. Yet an experience we seem set to have is not even an experience we are sure to have, let alone an experience we have had. And the mere fact that we seem, or are, set to have some experience surely makes us no better off now. So while the hedonist was right, her picture was incomplete. Our welfare is constituted entirely by enjoyable experience; and our welfare judgments track which things have proven themselves to us to reliably lead to such experience.

7.3 Wrapping up

The payoff for the hedonist, if they have followed me this far, is great. In showing that they can account for any intuition concerning the features of lives that seem relevant to the going-well and doing-well "dimensions" of well-being, I've shown that the hedonist can account for any intuition concerning the "objective" features of a life "external" to a subject's experience, as well as the "subjective" stances toward a subject's own life, that appear to impact the subject's well-being. As we have seen, the hedonist can claim that these intuitions reflect our genuine desires, which have been reinforced in accordance with how well they have led to enjoyable and unenjoyable experience. This means that these intuitions are tracking the features of lives that have proven themselves to us to reliably lead to enjoyable and unenjoyable experiences for the subjects of those lives. And this means that our intuitions concerning "objective," "external," and relevant "subjective" features of a life

are merely reflections of the fact that these features have proven themselves to us to reliably lead to enjoyable or unenjoyable experiences.

Importantly, notice that this picture is not one on which the hedonist is “debunking” our intuitions concerning the relevant objective, external, or subjective features of a life. The hedonist is not saying that while our intuitions point in this direction, they are simply misleading. On the contrary, she is saying that these intuitions are pointing us in the right direction, toward the things that have proven themselves to us to reliably lead to enjoyment. The only intuition that the hedonist has “debunked” here is Lin’s comparison intuition. And that, it should be noted, is only because that intuition is of a special kind. Specifically, it is an intuition concerning the well-being of lives that have concluded. We are justified in dismissing this intuition because we can be sure that our intuitions of this kind are leading us astray. Since we know, by stipulation, that lack of contact with reality will not lead to any unenjoyable experience, we are justified in distrusting this intuition, given that it merely reflects the fact that lack of contact with reality has proven itself to reliably lead to unenjoyable experience. In contrast, in most cases where the lives haven’t yet concluded, we can’t be sure that some feature will or will not lead to unenjoyable experience, and so our evaluative judgments retain their full weight. Still, these judgments do not tell us which things are basically good for us; only which things reliably lead to the thing that is basically good for us, or to enjoyment.

As a final point, notice that this picture plausibly explains the clash between our and the deceived subject’s evaluations of their well-being. Since we know that their partner is cheating on them, and that their friends are mocking them, we see their lives as including a number of features that have proven themselves to us to reliably lead to unpleasantness and that, since they are deceived, the subject cannot see. Indeed, it is precisely because they

cannot see these indicators that they take their lives to be a-ok. Notice that we would agree with the subject on the matter of whether they were happy: they certainly see themselves as getting what they genuinely desire. But our judgments on their welfare would diverge. That is because we know that their happiness is unlikely to last, and they do not. We know that they will not be getting what they genuinely desire—that their happiness is not reliable. They, of course, see things differently: they have a loving partner and loyal friends. They think that their happiness is reliable. But it is not.

Since deception has proven itself to us to be a reliable indicator of future unenjoyable experience, our evaluative intuitions reflect a general rule: if a subject is deceived, then she has some painful times ahead. It should then come as little surprise that, when we stipulate that she does not have painful times ahead, our intuitions point us in the wrong direction. Our general rule is stipulated to fail, and it was not formed in a way that allows it to recognize exceptions. We thus feel the residue of the rule's verdict, even when we know that the verdict leads us astray.

Our situation in the debate over the nature of well-being is not, then, perfectly analogous to the situation you face with Ann. In your case, the general rule—don't rely on Ann—is not quite as fixed, and your evaluative intuition is not quite as recalcitrant, as our general rules and evaluative intuitions about well-being. Nonetheless, the general idea is the same. Relying on the unreliable makes us uncomfortable—and, in most cases, rightly so. But sometimes, this aversion leads us only in the wrong direction.

7.4 Reliabilist hedonism vs. the genuine desire theory of welfare

The theory of well-being that I have defended in this chapter tells us that our many intuitions about the things that contribute to our well-being are explained by the fact that the

things have proven themselves to us to reliably lead to enjoyment. Since all of the things in life that seem good for us also reliably lead to enjoyment, this picture plausibly explains all of our intuitions about changes in well-being while also vindicating the hedonist's claim that all and only enjoyable experiences are the basic constituents of our well-being. This leaves the hedonist with a view on which well-being (broadly conceived) is a kind of reliable happiness. I will therefore call the view *reliabilist hedonism*. On this view, a subject is happy over some interval if and only if, because, and to the extent that they see themselves as getting what they genuinely desire, and are unhappy if and only if, because, and to the extent that they see themselves as getting what they're genuinely averse to. A subject's well-being is also constituted entirely by experience in which they see themselves as getting what they're genuinely attracted to; but, their welfare *judgments* track which things have proven themselves to the subject to reliably lead to such experience. Put differently: reliabilist hedonism says that the basic constituents of our welfare are the same as the basic constituents of our happiness—all and only enjoyable experiences—yet our welfare judgments are made distinct by the fact that they are predictions of *future* happiness.

One of the key claims I have argued for in this chapter is that reliabilist hedonism plausibly explains our many, perhaps seemingly inconsistent, intuitions about which things are good for us. In order to see the connection between this view and the genuine desire theory of well-being, it may be helpful for me to briefly explain how reliabilist hedonism makes sense of what I take to be a particularly important intuition here. (I will return to this point in the following chapter to explain it in much greater detail.) This is the intuition that has led many to conclude that well-being consists in, not just any enjoyment, but only enjoyment that is taken in "objectively good" things. The intuition, which I have discussed at length throughout this dissertation (chs. 2, 3, 4, 6, and 7), is that while enjoyment is no doubt

a key factor to how well our lives go, there is an important sense in which we only seem to be doing well provided we are enjoying the "right" things. After all, someone who gets a great deal of enjoyment from watching innocent people being tortured does not seem to be doing as well as someone who gets the same amount of enjoyment from acting virtuously. Nor does the drug addict seem to be doing as well as someone who gets an equal amount of enjoyment from their work as a doctor. This is explained, on the current account, by the fact that in each case the latter person's enjoyment is taken in something that has proven itself to us to reliably lead to further enjoyment; and the former's is taken in something that has proven itself to us to reliably lead to pain. Those who act virtuously or help others are not only more likely to be treated favorably by others, but are also likely to have resulting feelings of pride and fulfillment. And those who act nefariously or recklessly are not only more likely to be treated unfavorably by others, but are also likely to have resulting feelings of shame and regret. The reason we seem to be doing better when we enjoy the "right" things, then, is not that the "right" things are "objectively good" things, but that the "right" things are just those that reliably lead to further enjoyment, or to a greater amount of the good. Reliabilist hedonism thus plausibly explains what I take to be an important intuition concerning the nature of well-being.

Now recall what the genuine desire theory of well-being had to say about these cases. On that theory, the reason we seem to be doing better when we enjoy the "right" things is that these things are objectively good. More precisely, the view claims that enjoyment taken in objectively good things, or things for which there are (more) reasons for anyone capable of desiring the things to desire them, makes a greater contribution to a subject's well-being than an equal amount of enjoyment that is taken in things for which there are no (or fewer) such reasons. Since there are more reasons for anyone capable of desiring to act virtuously,

or help others, to do these things than there are reasons for anyone capable of desiring to torture innocent people, or take addicting drugs, to do these things, the theory implies that acting virtuously and helping others have a greater amount of objective worth than torturing innocent people and taking addicting drugs. That, the theory implies, is why those who enjoy the "right" things, or things like acting virtuously and helping others, are doing better than those who equally enjoy the "wrong" things, or things like torturing innocent people and taking addicting drugs.

In the previous chapter, I suggested that there might seem to be something wrong with this explanation. In particular, it might seem that this explanation is not sufficiently unifying. The genuine desire theory of well-being explains the great bulk of our intuitions about changes in well-being in terms of enjoyment, and then retreats to talk of objective goodness in order to settle the cases where enjoyment alone seems incapable of doing all of the explaining. A different way of pressing this potential objection may be to ask: why are there more reasons for anyone to desire some things than there are reasons for anyone to desire others? If the explanation is not ultimately rooted in enjoyment, the theory may seem decidedly less hedonistic, and indeed decidedly less plausible. After all, this would mean that there was some deeper fact, unrelated to enjoyment, that played some role in the explanation of which things are basically good for us. In that case, we might wonder why enjoyment was the focus of our theory in the first place.

Reliabilist hedonism helps to show that this objection is ultimately nothing to worry about. I think it is correct that, if our explanation as to why there are more reasons for us to desire some things rather than others was not ultimately rooted in enjoyment, the theory might lose at least some of its plausibility. But notice that reliabilist hedonism shows that our explanation here may in fact ultimately be rooted in enjoyment. More specifically, the view

claims that things seem good for us in accordance with how well they have led us, and so how well they seem likely to lead us, to enjoyment. This means that the greater the contribution a thing seems to make to our well-being, the greater the amount of enjoyment the thing has proven itself to us to lead to. And, of course, the fact that something will lead to enjoyment is a reason to desire it; and the fact that one thing will lead to more enjoyment than another is a further (or stronger) reason to desire the one thing, and is no reason at all to desire the other. It then follows that (other things being equal) we have more reasons to desire things that lead to a greater amount of enjoyment than we have reasons to desire things that lead to a lesser amount of enjoyment. And so, the reason that we have more reason to desire some things than we have reason to desire others is that some things lead to a greater amount of enjoyment than others. Our explanation is thus ultimately rooted in enjoyment, and the problem is dissolved.

It is crucial to note, however, that this way of dissolving the problem ultimately suggests that it is reliabilist hedonism, rather than the genuine desire theory of welfare, that offers the more fundamental explanation as to the nature of well-being. After all, if the objective worth posited by the genuine desire theory is ultimately to be explained in terms of enjoyment in the way reliabilist hedonism suggests, then the more fundamental explanation as to the nature of welfare lies with reliabilist hedonism. Since there seems to be no other, more plausible way of further analyzing the nature of objective worth, I conclude that reliabilist hedonism offers the most complete and fundamental explanation of the nature of well-being, and that it should therefore be preferred over not just the genuine desire theory but all other alternatives.

7.5 Looking forward

In the following chapter, I want to show that reliabilist hedonism provides the hedonist with a plausible way of accounting for several objections that are often taken to significantly undermine her view. These are objections concerning objective goods, external goods, and the nature of pleasure. Though these objections are often taken to provide us with sufficient reason to reject hedonism, I will show that, given reliabilist hedonism, they do not.

8 Hedonism Revived

Recall that, on

> *Hedonism*: all and only enjoyable (unpleasant) experiences are basically good (bad) for a subject; and an enjoyable (unpleasant) experience contributes to a subject's well-being (ill-being) just to the extent that the experience has a greater intensity or duration.[84]

Despite its simplicity, hedonism about well-being has at least one crucial advantage over competing theories: it naturally accounts for the near-universal intuition that enjoyment is a good thing. While some deny that it is always good, and many deny that it is the only good, it is very difficult to deny that it is *sometimes* good. In fact, it seems difficult to deny that at least some enjoyment is *required* for a life to count as good. The task for the hedonist is to build on these shared intuitions in a way that will convince others that enjoyment is all we need in our theory of welfare—or, convince them that all and only enjoyable experiences are basically good for us.

Unfortunately for the hedonist, there are a number of convincing, and apparently fatal, objections to her view. For one thing, there are a number of cases where the veracity of

[84] Again, I follow Crisp (2006) in formulating hedonism in terms of enjoyment, rather than pleasure, for reasons he notes. I also use "unpleasant," rather than "painful," for parallel reasons. As the title of this chapter indicates, I hope here to build on Crisp's efforts in the article on which that chapter of his book is based to "reconsider" hedonism.

our intuitions seems incompatible with the truth of hedonism. Here there have been two particularly influential types of case. First, a problem for hedonism arises from the common intuition that we only seem to be doing well provided we are enjoying the "right" things—things that are most often described as being "objectively good."[85] Enjoyment derived from, for instance, torturing others, engaging in bestiality, and taking addictive drugs, does not seem to contribute as much to one's welfare as enjoyment of the same intensity and duration derived from acting virtuously, loving one's partner, and doing meaningful work. Yet, if all that matters to one's well-being is the extent to which one enjoys oneself, the particular things that are enjoyed should not matter to one's well-being. Since the things we enjoy do seem relevant to how well we are doing, the veracity of our intuitions here seems inconsistent with the truth of hedonism. We can call this *the objectivity objection*.

Second, a related worry for hedonism arises from the common thought that things "external" to our experience can influence our well-being. Indeed, perhaps the most common way of objecting to hedonism is by appealing, as Shelly Kagan does, to cases where a person's life seems to them to be as great as it could be, yet they are in fact being completely deceived: their seemingly loving partner is cheating on them, their seemingly loving children only pretend to like them for their money, and their seemingly successful business is set to go bankrupt due to a corrupt business partner.[86] Intuitively, this person does not seem to be doing as well as they would be if they were not being deceived in these

[85] This problem is mentioned as early as Aristotle (2000, book X), and is perhaps most notably presented by G. E. Moore (1903, ch. 3). More recently, it is also frequently invoked by those who wish to motivate the hybrid view (which I have discussed in chapter 2, and will discuss again in section 2 of this chapter) on which well-being consists in "enjoying the good."

[86] Kagan 1994 (p. 311). Also relevant here are cases involving Nozick's "experience machine," which are meant to make essentially the same point as Kagan's, which is that deception in itself can make a life worse (1974, pp. 42-45). Since I have discussed cases involving the experience machine in chapter 7, and am in any case interested in the arguments Kagan provides in light of his example, I will focus on Kagan's case here.

ways. Yet, if our enjoyable experiences are all that matter to our well-being—as hedonism claims—it follows that this person cannot be worse off in virtue of being deceived, since this deception has no effect on their enjoyable experiences. A similar worry, whose most notable presentation comes from David Velleman, arises from the intuition that a life that starts poorly and trends upwards seems better than a life that starts well and trends downwards, even if both lives contain the same amounts of enjoyment and unpleasantness.[87] If hedonism is correct, however, then the "shape" a life takes could not in itself affect how well a life goes, since this shaping in itself has no bearing on one's enjoyable or unpleasant experiences. Again, the veracity of our intuitions here seems inconsistent with the truth of hedonism. We can call this *the externality objection*.

For another thing, hedonism seems to face an important theoretical issue. It is no doubt important for the hedonist to provide a plausible account of what enjoyment, or the thing that they take to be the sole basic good, consists in.[88] Here there are two leading approaches: phenomenological theories, which take enjoyment to consist in a distinct phenomenology, such as a warm glow; and attitudinal theories, which take enjoyment to consist in a special sort of positive attitude taken toward some object.[89] Yet, each of these theories is said to face significant issues. Against phenomenological theories, notice that the enjoyment involved in, for example, taking a warm shower after a freezing day, reading a great novel, and spending time with a loved one, seem rather different in nature. There does

---

[87] Velleman 1991.

[88] This debate has more often been framed in terms of pleasure, but, for reasons I will provide in section 4, I believe it is best to instead focus on enjoyment.

[89] Proponents of particular phenomenological approaches include Bramble (2013), Crisp (2006), and Kagan (1992). Proponents of particular attitudinal approaches include Feldman (2004) and Heathwood (2006, 2007). As I've mentioned, these two types of theory are also sometimes known as "internalist" and "externalist" theories of pleasure, respectively. See Sumner 1996 (pp. 87-92).

not appear to be any unifying phenomenological feature of these experiences of enjoyment that can explain what makes the experiences enjoyable, which suggests that phenomenological theories are false.[90] On the other hand, against attitudinal theories, notice that enjoyment does not typically seem to be mediated by any attitude on the part of the person who experiences it: when they enjoy something, such as the feeling of a warm shower, they never stop to form an attitude about the thing before they enjoy it—their enjoyment instead seems more immediate. More generally, it is difficult to see exactly what attitude would be required for enjoyment, or what relation this attitude would have to share with its object to make the experience an enjoyable one. We can call the objections to these theories of enjoyment, taken together, *the enjoyment problem*.

In this paper, I want to show that none of these objections undermine hedonism. On the view I have called reliabilist hedonism, we can account for all of our intuitions about changes in well-being, as well as the apparent theoretical issues hedonism seems to face, all while retaining the claim that all and only enjoyable experiences are basically good for us. In the following section, I will briefly review this version of hedonism. Then, in sections 2-4, I will show how it provides the hedonist with plausible responses to the objectivity and externality objections, as well as the enjoyment problem, respectively. I'll conclude that reliabilist hedonism constitutes a revival of hedonism.

[90] As I've mentioned, this is known as the "heterogeneity" problem for phenomenological theories of pleasure.

8.1 Reliabilist hedonism

To solve these problems for hedonism, it will be helpful for me to once again place the view in what I take to be its proper context. This will consist of reviewing three claims about the nature of enjoyment and its relation to our evaluative intuitions.

The first and perhaps most controversial claim is that enjoyable experiences are partially constituted by some of our intrinsic desires. More specifically: sometimes when we intrinsically desire things, we are genuinely attracted to the things—we are not merely disposed to get them, but are enthused or excited by them. We have called desire of this sort *genuine* desire (chs. 3, 4, 7). As I've mentioned, I think that enjoyable experience *just is* experience in which we see ourselves as getting what we genuinely desire. That is, it is just experience in which we take something we are enthusiastic or excited about to obtain. This is evidenced by the fact that any enjoyable experience also seems to be experience in which we take ourselves to be getting something we are genuinely attracted to. Activities as diverse as eating, reading, working, and relaxing are all candidates for being enjoyable, and what seems to make them enjoyable is our genuine attraction to doing them. Note also that this approach can plausibly explain the nature of the two most commonly discussed types of pleasure: attitudinal pleasure and sensory pleasure.[91] Attitudinal pleasure is pleasure about some state of affairs, such as the fact that one's favorite team has won. Sensory pleasure is pleasure experienced in relation to some pleasurable sensation, such as the feeling of a warm shower after a freezing day. On the picture I'm proposing, an experience of attitudinal pleasure in some state of affairs is just an experience in which one genuinely desires that state affairs and takes that state of affairs to obtain; and an experience of sensory pleasure is

[91] See Feldman 2004 (pp. 55-57), Heathwood 2007 (p. 28), and Lin 2018 (p. 2).

just an experience in which one feels a particular sensation and genuinely desires to be feeling that sensation. More generally: any enjoyable experience is just an experience in which we take ourselves to be getting something that we are genuinely attracted to.

The second claim is that our evaluative intuitions are in part reflections of our genuine desires and aversions. Thus, for example, when I have the intuition that pain is bad for a subject, this is in part a reflection of my aversion to pain. Notice that this takes no stand on the precise nature of this reflection: the claim is just that our particular genuine attractions and aversions in some way help to determine the contents of our evaluative intuitions.

Finally, the third claim is that we have come to possess our desires and aversions by way of a particular process—one endorsed by Richard Brandt and Peter Railton, among others.[92] The idea here is that we have been conditioned, via a process of positive and negative reinforcement, to desire things that have led us to enjoyment, and to be averse to things that have led us to unpleasantness. The process works roughly as follows. When a desire leads us to enjoyable experience, it is positively reinforced, making us more likely to retain it and to act on it in the future; and when a desire leads us to unenjoyable experience, it is negatively reinforced, making us less likely to retain it and to act on it in the future. Now, as the so-called paradox of hedonism shows us, those who aim directly at enjoyable experience are often less likely to achieve it than those who aim at things that tend to lead to enjoyable experience. The enjoyable experiences associated with a loving relationship, for instance, are not attainable by someone who intrinsically desires only those experiences, since any relationship formed purely on the basis of that desire will not count as loving. It

[92] See Brandt 1979 (p. 100) and Railton 1989 (p. 167). See also Arpaly and Schroeder 2014 (ch. 6) and Schroeder 2004 (chs. 2, 3, and 4).

then seems that the conditionalization process would leave us with intrinsic desires for things that were not themselves, but that themselves tended to lead to, enjoyable experiences. All of our intrinsic desires for things would then be explainable in terms of their proven tendency to lead to enjoyment. Besides having a good deal of intuitive plausibility, this picture of how our desires are conditioned also fits well with the earlier claim that there is a constitutive connection between enjoyment and genuine desire. That's because, if genuine desire is constitutive of enjoyable experience, it will be no wonder that our genuine desires are conditioned in accordance with how well they lead to enjoyment. Our genuine desires aim at enjoyable experience, and are reinforced in accordance with how well they attain it.

Now here is the resulting account. Our genuine desires have been conditioned in a way that ensures that their objects reliably lead to enjoyment. On the picture I'm proposing, our intuitive welfare judgments reflect our genuine desires, such that these judgments are tracking which things have proven themselves to reliably lead to enjoyment, and the extent to which they have. That is, when we say that things like being respected by our peers, having a loving partner and family, and the like, are good for us, we are indicating that these things have proven themselves to us to reliably lead to enjoyment. Since all of the things that seem good for us are also things that seem to reliably lead to enjoyment, this picture plausibly explains our intuitions about the great variety of things that seem good for us, while also vindicating the hedonist's claim that all and only enjoyable experiences are basically good for us. This leaves the hedonist with a picture on which well-being (broadly conceived) is a kind of reliable happiness. I will therefore call the view

Reliabilist hedonism: all and only enjoyable (unpleasant) experiences are basically good (bad) for a subject; an enjoyable (unpleasant) experience contributes to a subject's well-being (ill-being) just to the extent that the experience has a greater intensity or duration; and an evaluator's intuitions about the goodness (badness) of particular things for a subject track the extent to which the things have proven themselves to the evaluator to reliably lead to enjoyment (unpleasantness).

This formulation cuts to the view's core: the first two claims are just hedonism, and the third claim is the upshot of the three further claims I have just argued for.

With the proposal in place, let's return to the worries for hedonism outlined above.

8.2 The objectivity objection

It has become commonplace in the literature on well-being to object to hedonism on the grounds that not all enjoyment of a given intensity and duration is equally good for us—or, to appeal to the objectivity objection. This objection can be seen in the claim that enjoyment derived from, for instance, taking addictive drugs does not seem as good for one as enjoyment of the same intensity and duration derived from one's work as a doctor. It is also the same objection that moved Mill to famously argue that there were not only differences in quantity between pleasures, but also differences in quality.[93] This argumentative move allowed him to claim that pleasures requiring the use of one's "higher faculties" or intellect are more valuable than pleasures that are achievable by the "lower" animals, even if the pleasures are equal in terms of intensity and duration. On this picture, the "higher" pleasures

[93] Mill 1863 (ch. 2).

possess an "intrinsic superiority" over the "lower" pleasures; so much so, in fact, that Mill famously wrote: "It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates satisfied than a fool satisfied" (1863, ch. 2). The intrinsic superiority of the pleasures experienced by human beings is then enough to make a life containing very few of these pleasures better than a life full of pleasures experienced by animals like pigs.

Of course, it is not entirely clear how Mill could justify his contention here while also leaving his account a proper form of hedonism. After all, as I've mentioned in the previous two chapters, if enjoyment is the sole basic good, we should expect that any explanation of differences in goodness and badness between particular things should ultimately appeal to enjoyment.[94] That is, if enjoyment is the sole basic good, then the ultimate explanation of any difference in the goodness of two experiences should presumably appeal just to a difference in enjoyment. Yet any explanation of these differences that appeals to "intrinsic" or "objective" goodness, or to further features of the experiences, such as whether they are aesthetically or morally better, seems to sneak into one's account of goodness something that goes beyond mere enjoyment. It seems inconsistent with a theory like hedonism to claim that some feature of experiences, beyond mere enjoyableness—understood just in terms of intensity and duration—should play any role in the explanation of what is basically good for one, and this seems to present a problem for Mill's account.

But regardless of whether this presents an insurmountable hurdle for Mill's view, it is worth noting that the underlying motivation for Mill's introduction of qualitative aspects of pleasures into his account has also motivated a rather different approach to well-being. As I've mentioned, many of us share the intuition that experiences of enjoyment of a given

[94] Moore (1903, ch. 3) makes a similar point.

intensity and duration do not always seem equally good for us. When we hold the intensity and duration fixed, the enjoyment involved in, say, torturing others, or engaging in bestiality, does not seem as good for us as the enjoyment involved in acting virtuously, or connecting with one's partner. While this led Mill to claim that there is a qualitative dimension to pleasures, all of which nevertheless remain good for one, it has led others to claim that there is a proper subset of pleasures—pleasures taken in the "right" things, or in "objective goods"—that are basically good for one, and that pleasures outside of this subset fail to contribute to one's well-being. These latter views, which are often seen as "hybrids" of hedonism and objective list accounts, and which we have discussed at length in chapter 2, can be captured by the slogan that well-being consists in "enjoying the good." The primary motivation for adopting these views, which is also the primary motivation their proponents provide for rejecting hedonism—namely, that enjoyment taken in some things seems better for us than enjoyment of the same intensity and duration taken in other things—is what we have called the objectivity objection.

How are we to understand which things are the "right" ones, or are "objectively good," and thus which experiences of enjoyment contribute to our well-being, on these views? More often than not, proponents of these views are quiet on this question. Nonetheless, there is a clear intuitive understanding of which things will count as objectively good, and which things will not. Friendship, achievement, and autonomy will no doubt be included on the list of objective goods, while pain, frustration, and ignorance might be included on the opposing list. What *makes* these things objectively good or bad? The only plausible answer available here seems to be: they *just are*. Just as the hedonist can provide no answer to the question what makes enjoyment good for us other than that it just is, so proponents of these views may be forced to concede that there is no further explanation

they can provide as to what makes some things objectively good. Perhaps the most that can be said is that they just are.

Regardless of how plausible we take this response to be, I believe we can do better. In fact, I believe we can show that the very idea of something (other than enjoyment) being "objectively" good arises from a confusion. The reason some things seem "objectively" good is not that they just have this goodness, but is instead that they are conducive to enjoyment—or, that they are *conducive to* the *sole basic good*. Reliabilist hedonism provides the basis for this explanation.

To see this, notice first that reliabilist hedonism offers the hedonist a plausible response to the objectivity objection. This objection claims that not all experiences of enjoyment of the same intensity and duration seem equally good for us. And it is problematic for the hedonist, of course, because she claims that our level of well-being is determined entirely by the intensity and duration of our enjoyable experiences. Given reliabilist hedonism, however, it seems clear that the objectivity objection need not be a problem for the hedonist. This is because she has an alternative explanation as to why some enjoyable experiences seem better than others despite being equal in terms of intensity and duration. On this explanation, these experiences seem better, not because the enjoyments they contain are taken in more "objectively" valuable things, but because the enjoyments they contain are taken in things that have proven themselves to us to more reliably lead to enjoyment. That is, the enjoyment involved in, say, torturing others, engaging in bestiality, and taking addictive drugs seems worse to us than the enjoyment of the same intensity and duration involved in acting virtuously, loving one's partner, and doing meaningful work, not because the latter things have more objective value than the former things, but because the latter things have proven themselves to us to more reliably lead to greater enjoyment.

How plausible is this alternative explanation offered by the hedonist? As I see it, it is at least more plausible than the appeal to objective goodness. After all, proponents of any such appeal must eventually provide an explanation as to what this objective goodness consists in, and how we are to understand its mysterious presence in some things and not others. It seems far more natural to instead appeal to a thing's conduciveness to further enjoyment, which is plausibly what our desires are ultimately aiming at, in explaining why some things seem to have a type of value that goes beyond any attitudes that are currently had towards the things. Rather than positing any mysterious properties that stand in need of a more complete explanation, this view plausibly explains the nature of the good by appealing merely to basic psychological phenomena, such as desire and enjoyment, allowing for a plausible naturalistic reduction of the good.

Moreover, reliabilist hedonism can plausibly explain the divergence in judgments about the goodness and badness of particular things between normal adult human beings and people with backwards values. Consider, for instance, Bea. Bea had a disturbing childhood, and as a result now enjoys nothing more than harming and killing others. In fact, these are the only things she enjoys. Whereas normal adult human beings would be likely to see harming and killing others as bad, and to have the intuition that these things would not make them any better off, Bea sees them as good, and is more than happy to report her intuition that they would make her much better off. After all, these are the only things she enjoys, and she enjoys them very much. Now, when Bea reports that these things contribute to her well-being, and that virtuous action and the like would not, we might think that she simply betrays her complete lack of grasp of the concept. However, I think a more plausible explanation is that, while she grasps the concept perfectly well, the things she has been conditioned to see as reliably leading to enjoyment are very different from the things we

have been conditioned to see as reliably leading to enjoyment. It is not as if she fails to understand what makes something good for someone; rather, it's that her genuine desires, which determine which things an evaluator sees as being good or bad for a subject, are radically different from our own. Reliabilist hedonism thus plausibly explains this divergence in judgments. Of course, regardless of how we explain Bea's judgments, proponents of other theories of welfare are still free to claim that these judgments are simply *wrong*. But I think that would be a mistake. As I see it, it seems plausible that Bea is not simply mistaken about the extent to which harming and killing others is good for her. If we suppose that her psychology is so constituted that her genuine desire to harm and kill others will never be overturned or weakened in the future, and her situation is such that she will never be without people to harm or kill and will never face repercussions for doing so, it seems to me that Bea might be just as correct to claim that harming and killing others contributes a great deal to her welfare as we are correct to claim that (say) acting virtuously or loving one's partner contribute a great deal to our welfare. (Alternatively: suppose Bea is thriving in a violent society—perhaps one that is inhabited by more violent types of beings than us—where psychologies such as Bea's are the norm.) Yet if our welfare partly depends on the objective worth of the features of our lives or of the things we enjoy, then this could not be the case. It could not be that Bea's well-being is improved by hurting and killing others to just the same extent that our well-being is improved by acting virtuously, because hurting and killing others (it would presumably be said) has far less objective worth than acting virtuously. Provided hurting and killing others would continue to lead her to enjoyment, however, and would continue to be positively reinforced in the future, I think reliabilist hedonism correctly implies that Bea's well-being could be improved by hurting and killing

others to just the same extent that our well-being is improved by acting virtuously.[95] If this is right, it represents a further reason to prefer reliabilist hedonism to theories of well-being that appeal to objective goodness.

It seems clear, in any case, that reliabilist hedonism allows the hedonist to avoid the objectivity objection in a way that is plausible, and that does not undermine the theory's status as a proper form of hedonism. The ultimate explanation as to why some things seem more objectively valuable than others, and thus why experiences of enjoyment of the same intensity and duration do not always seem equally good for us, is that some things have proven themselves to lead more reliably to enjoyment than others, and thus enjoyment taken in some things has proven itself to more reliably lead to more enjoyment than enjoyment of the same intensity and duration taken in others. Since this explanation appeals only to enjoyment in accounting for a perceived difference in value, the hedonist can adopt it without undermining her commitment to hedonism. The objectivity objection thus presents no worry to the hedonist.

---

[95] This might lead us to question whether the view can also plausibly explain cases where there is a divergence in judgments about the goodness or badness of particular things between normal adult human beings whose everyday values or tastes differ. It might seem that it cannot. Suppose, for instance, that Ann very much likes durian, and thus sees eating it as good. By contrast, if your palate is anything like mine, you'll likely find the taste and smell of the fruit putrid, and thus see eating it as bad. Surely, though, we do not have the intuition that eating durian everyday would be bad *for Ann*, nor would she have the intuition that doing so would be good *for us*. If our evaluative intuitions about the goodness and badness of things are determined just by how our own desires and aversions have been conditioned, though, our intuitions presumably should not change when we consider the goodness or badness of those things for others. So why do they? The answer lies in our awareness of the fact that Ann enjoys eating durian. If we were not aware of this fact, we would presumably not have the intuition that eating durian everyday would be good for her. And since we are aware of it, the question of whether eating durian everyday would be good for her is to us roughly equivalent to the question of whether eating something she enjoys (but we do not) every day would be good for her. And crucially, unlike eating durian every day, eating what one enjoys everyday seems just as good to us as it seems good for others. So our intuition about whether durian is good for Ann changes when we learn that she enjoys eating durian because we are in a sense no longer evaluating the goodness of eating durian, and are instead evaluating the goodness of eating what one enjoys. If we had more information—about durian's nutritional profile, for instance—then that too would be a part of what we were evaluating. This illustrates that the more information we have about either the thing we're evaluating or the subject who would get it, the more complex our evaluative intuitions will grow.

8.3 The externality objection

Consider next the externality objection. This is the objection that things "external" to our experience, such as whether we are deceived, or whether our life takes one "shape" rather than another, can impact our well-being. Since hedonism claims that well-being is just a matter of enjoyable experiences, and deception and the shapes of our lives need not influence our enjoyable experiences, our intuition that such deception and shaping can nevertheless make a subject worse off seems to conflict with the truth of hedonism.

This has led some, such as Shelly Kagan and David Velleman, to go so far as to claim that our concept of well-being alone is insufficient to account for our intuitions surrounding a person's good. Kagan in particular argues for a distinction between being well off, or having a high level of *welfare*, and living a life that goes well, or having a good *life*.[96] This would allow us to claim that, for instance, while the completely deceived businesswoman was doing well, or had a high level of welfare, her *life* (as a whole) was not going nearly so well. According to Kagan, theories that focus just on our experiences, like hedonism, might be plausible as theories of what it is to be well off, but cannot plausibly account for what it is to have a good life. This is, in effect, because a good life requires that things external to our experiences, such as whether we are deceived, also go well for us. What it is to have a good life will thus likely be a more objective matter than what it is to be well off, in that the former may encompass features of our lives that we are entirely unaware of.

[96] Kagan 1994 (pp. 319-321).

I believe that this proposed distinction rests upon a mistake, however. This is the mistake of taking the goodness of our experiences to be incapable of explaining the goodness of things external to our experiences. That is, it is the mistake of taking the goodness of things external to our experiences to require a concept that encompasses more than just the goodness of our experiences. This is a mistake because, on reliabilist hedonism, the goodness of our experiences can plausibly explain the goodness of things outside of our experiences, leaving our experiences as the more basic good. On this theory, all and only enjoyable (unpleasant) experiences are basically good (bad) for us, and our evaluative intuitions about which things are good (bad) for us track which things have proven themselves to us to reliably lead to enjoyable (unpleasant) experience. When we say that things external to our experience are good (bad) for us, then, we are saying that these things have proven themselves to us to reliably lead to enjoyable (unpleasant) experience. This means that subjects of deception, such as the deceived businesswoman, seem to us to be made worse off by that deception because deception is to us a reliable indicator of future unpleasant experience. Indeed, what makes the deception she suffers seem particularly bad is that she will likely soon be robbed of many of the things in her life that we take to reliably lead to enjoyment. Her relationships with her kids, partner, and business partner will never be the same. She seems set to have a bunch of extremely unpleasant experiences ahead of her. On the other hand, if we add the further assumption that she has many other sources of enjoyment in her life, and that she will be able to rely on these while recovering from losing the others, her story, while of course still terrible, certainly seems less tragic. What all of this suggests is that our intuitions are tracking the features of her life that have proven themselves to us to reliably lead to enjoyable experience. Since reliabilist hedonism explains this fact while maintaining that all and only enjoyable experiences are basically good for us, it

follows that this theory plausibly explains the goodness of things external to our experience in terms of the (basic) goodness of our enjoyable experiences.

Nonetheless, it may at this point be objected: your explanation in the case discussed is plausible enough, but it will not be if we change just one feature of the case. Suppose instead that the deceived businesswoman's life has already ended, just before all of the terrible deception would have been revealed to her. According to reliabilist hedonism, the fact that she was deceived seems to make her worse off because this deception is a reliable indicator of future unpleasantness. But we know that she won't face any future unpleasantness: after all, her life has already ended. So, reliabilist hedonism explains our intuition that she is worse off in a way that fails to reconcile it with something we surely know. And this undermines the view.

Contrary to the objector's suggestion, however, I do not believe that this undermines reliabilist hedonism. On the contrary, I believe that it shows something that we should simply accept: that our intuitions about changes in the well-being of those whose lives have already ended are misguided. These intuitions help lead us, and help us to lead others, to things that we enjoy, and thus to a greater amount of enjoyment overall. If that is right, then it should be no wonder that when we are asked about changes in the well-being of those whose lives have already ended, our intuitions misfire. They retain their force, and retain their legitimacy insofar as they are leading us to things that have proven themselves to us to lead to more enjoyment overall, but they ultimately misguide us, because the people they concern are no longer capable of enjoyment. I take this explanation of our intuitions about posthumous benefits and harms to be at least as plausible as any alternative explanation, especially given the plausibility of the overall picture of which it is a part.

In a way that is similar to Kagan, Velleman argues that we must distinguish between a person's momentary and lifetime well-being, and that a person's lifetime well-being is not a simple aggregation of their momentary well-being.[97] In support of this, Velleman appeals to cases where the "shape" a life takes itself seems to affect how well the life goes. The idea here is that the "shape-of-a-life," such as beginning well and steadily trending downward from there, or vice versa, can itself influence how well the life actually goes. Intuitively, a life that starts well and gets worse, ending with unenjoyable or mediocre experiences, seems worse than a life that starts poorly and gets better, ending with enjoyable experiences, even if those lives contain equal amounts of enjoyment and unpleasantness in total. This suggests that enjoyment, though a plausible candidate for constituting momentary well-being, cannot fully account for lifetime well-being. More generally, Velleman thinks, momentary well-being cannot fully account for lifetime well-being, in that lifetime well-being cannot be a simple aggregation of momentary well-being. Rather, the narrative relations between the events making up one's momentary well-being are also an essential element in determining one's lifetime well-being. Lifetime well-being can only be understood, then, in terms of both momentary well-being, and the narrative significance of the overall trend one's momentary well-being displays over time.

Again, however, I think the most plausible explanation here is much simpler than the one Velleman provides. On reliabilist hedonism, the reason we see the improving life as better than the worsening life when each contains the same amount of enjoyment is that the improving life contains more features that have proven themselves to us to reliably lead to enjoyment than the worsening life does. Notice that the recipe for providing a case that will

[97] Velleman 1991 (pp. 53-55).

pump the intuition that the "shape-of-a-life" matters is to compare a life that begins with very few reliable indicators of future enjoyment and improves to contain many reliable indicators of future enjoyment with a life that begins with many reliable indicators of future enjoyment and worsens to include very few reliable indicators of future enjoyment, and to then stipulate that the lives contain the same amount of enjoyment in total. Indeed, the things that intuitively would constitute a life's improvement all seem to be things that have proven themselves to us to be reliable indicators of future enjoyment: going from, say, impoverishment to riches, addiction to sobriety, or a life of crime to a life of integrity, in each case seems to entail going from something that has proven itself to us to reliably lead to unpleasantness to something that has proven itself to us to reliably lead to enjoyment. We need not, then, posit more than one concept concerning a person's good, or make claims about the narrative relations between instances of goodness of one type, to make sense of the intuitive difference between the goodness of lives that become better over time and those that become worse over time, despite containing equal amounts of enjoyment. Our intuitions here are tracking the features of the lives that have proven themselves to us to be reliable indicators of future enjoyment. Since lives that become better also come to contain more features that have proven themselves to us to be reliable indicators of future enjoyment, and lives that become worse also come to contain fewer features that have proven themselves to us to be reliable indicators of future enjoyment, we have the intuition that the improving lives are better than the worsening lives, even though the lives in fact contain equal amounts of enjoyment and are thus equal in well-being to that point in their lives. As I see it, this explanation of our intuitions about the significance of a life's shape is at least as plausible as Velleman's, and it explains this phenomenon while consistently maintaining that all and only enjoyable experiences are basically good for us.

Of course, it may again be objected that, while my explanation here is plausible enough in cases where the lives we are considering have not ended, it becomes far less plausible when we consider lives that have concluded. After all, we do not just have the intuition that improving lives are better than equally enjoyable but worsening lives when the lives are set to continue: it still seems that a life that begins poorly and trends upwards until death is better than a life that begins well and trends downwards until death, even if the lives contain the same amount of enjoyment. Much like in the case of deception, however, I believe that our intuitions here can be explained in terms of their being simply misguided. These intuitions help lead us, and help us to lead others, to things that reliably lead to enjoyment, and thus to a greater amount of enjoyment in the future. In that case, when we ask about a person's welfare, and then stipulate that the person's life is over, it should be no wonder that our intuitions misfire. We see the life that has improved as including more features that have proven themselves to us to reliably lead to enjoyment at its end than we see the worsened life as including at its end. We thus have the intuition that the improved life is better than the worsened life. However, since we know that both lives contained equal amounts of enjoyment and unpleasantness, and that the amounts of enjoyment and unpleasantness the lives contained will not change in the future, we know that our intuition about the lives is misguided. The improved life was not in fact better than the worsened life, but rather, the lives were equally good. If the lives were to continue, however, we would not yet be free to conclude this. Again, I take this explanation of our intuitions surrounding the significance of a life's shape to be at least as plausible as any alternative explanation, including Velleman's.

I conclude, therefore, that reliabilist hedonism plausibly accounts for the externality objection. Things external to our experiences sometimes seem good for us because they

have proven themselves to us to be reliable indicators of future enjoyment, and they seem this way to us even when we know that future enjoyment is impossible.

8.4 The enjoyment problem

Consider next the enjoyment problem. This problem arises as a result of each of the two leading approaches to the nature of enjoyment, attitudinal and phenomenological theories, seeming to run into a distinct problem. According to phenomenological theories, enjoyment consists in a distinct kind of phenomenology, such as a warm glow. Yet the enjoyment involved in, say, reading a gripping novel, and that involved in taking a warm shower after a freezing day do not seem to share any particular phenomenological feature that might unify these experiences as enjoyable. Their phenomenology seems entirely distinct, which suggests that phenomenological theories of enjoyment cannot be made to fit with our intuitions. On the other hand, attitudinal theories claim that enjoyment consists in a certain sort of pro-attitude taken toward some object. Yet it is not clear what sort of attitude could play this role, or how the attitude must be related to one of its objects in order to play that role. Without a plausible account of the nature of this connection, attitudinal theories seem hopeless.

As I see it, however, the matter is not so dire. It's worth mentioning that this debate between phenomenological and attitudinal theories is usually framed in terms of pleasure, rather than enjoyment. This, I think, has confused the debate, since there are at least two commonly recognized types of pleasure: attitudinal and sensory. Attitudinal pleasure, recall, is the sort of pleasure one experiences when one takes pleasure "in" some fact, such as the fact that one's favorite team has won. The pleasure here is always related in this way to some state of affairs: attitudinal pleasure requires that there is some state of affairs in which

pleasure is taken. Sensory pleasure, in contrast, is the sort of pleasure one experiences when one has a pleasurable sensation or feeling, such as the feeling of taking a warm shower after a freezing day. Importantly, notice that it seems plausible that the attitudinal pleasure involved in, say, taking pleasure in the fact that one's favorite team has won may share nothing in common with the sensory pleasure involved in, say, taking a warm shower after a freezing day. That is, the phenomenology of these two types of experience may seem to share nothing in common, or at least nothing that could unify the experiences as enjoyable. Thus, by focusing on these two distinct types of pleasure, we find a reason to doubt the claim that pleasure consists in a distinct kind of phenomenology. Moreover, notice that, in light of this difference in the phenomenology of these two types of pleasure, it will likely be difficult to identify any particular attitude of the person who experiences them that might unify them as pleasures. That is, if what it is like to take pleasure in reading a gripping novel shares so little with what it is like to take a warm shower after a freezing day, then it will be difficult to plausibly conclude that there is nevertheless an attitude on the part of the person who experiences these pleasures that unifies them as pleasures. After all, they are so phenomenologically distinct, why think there is some attitude that unifies them?

We have thus made both attitudinal theories, which seem best suited to explain attitudinal pleasure, and phenomenological theories, which seem best suited to explain sensory pleasure, seem implausible in light of their inability to plausibly explain both types of pleasure. Yet, if we instead focus on the broader notion of enjoyment, it seems to me that this problem is easily resolved. Notice that experiences of attitudinal pleasure and experiences of sensory pleasure are, to the extent that they are so pleasurable, both enjoyable. Enjoyment thus encompasses both of these types of pleasure, and can also encompass other types of experiences, such as "flow" states, or states where one is entirely

absorbed by some activity, which are plausibly thought of as enjoyable. What do all of these experiences have in common? As I have already mentioned, I think that what unifies these states as enjoyable is that they all involve the subject of the experiences taking themselves to be getting something that they genuinely desire. That is, enjoyable experiences all involve that those experiencing them take themselves to be getting something that they are genuinely attracted to. This, I think, is the most plausible suggestion as to what enjoyment consists in: it consists in taking oneself to be getting what one genuinely desires.

If that is right, then we can also see that the enjoyment problem, and thus the debate between phenomenological and attitudinal theories, rests on a confusion. Recall that genuine desire is the sort of desire we have when we are genuinely attracted to or enthused by something. And, of course, there is a way that it feels to be genuinely attracted to or enthused by something—there is a distinct phenomenology inherent to this sort of attitude—as well as a way that it feels to see oneself as getting such a thing. Thus, if enjoyable experience just is experience in which we take ourselves to be getting something we genuinely desire, it follows that enjoyment consists in both a certain sort of pro-attitude—genuine desire—and a certain phenomenology—the way it feels to see oneself as getting what one is genuinely attracted to.[98] Accordingly, the assumption that enjoyment or pleasure must consist in just one of these things is mistaken. I hope to have shown here that the claim that enjoyment or pleasure *does* consist in just one of these things is mistaken as well.

Since reliabilist hedonism claims that enjoyable experience just is experience in which we take ourselves to be getting what we genuinely desire, and this conception of enjoyment

[98] This possibility is also explored by Lin (2018).

does not face the enjoyment problem, it follows that reliabilist hedonism does not face the enjoyment problem. In fact, the view plausibly explains how this apparent problem arose in the first place.

8.5 Conclusion

In this chapter, I have introduced and explored some of the implications of a brand of hedonism that I have called reliabilist hedonism. As we have seen, this view provides the hedonist with plausible responses to a number of significant objections that are often thought to undermine her view. The intuitions that things "external" to our experience, such whether we are deceived or our life takes a certain shape, and that the "objective" goodness of the things we enjoy, such as that had by acting virtuously and lacked by taking addictive drugs, seem relevant to our well-being can be explained by the reliabilist hedonist by appeal to the claim that our intuitions here track which things have proven themselves to us to reliably lead to further enjoyment or unpleasantness. The apparent relevance of these things to our well-being is thus explainable entirely in terms of enjoyment, and can thus be explained by the hedonist in a way that does not threaten her commitment to hedonism. Of course, these things also seem relevant to the quality of lives that have concluded, but we are justified in distrusting our intuitions about these lives, since we know that these intuitions attempt to track future enjoyment and unpleasantness, and that the subjects they concern are no longer capable of enjoyment or unpleasantness. Moreover, reliabilist hedonism solves the enjoyment problem. If enjoyment just is experience in which we take ourselves to be getting what we genuinely desire, then enjoyment will be constituted both by a certain attitude, namely genuine desire, and a certain phenomenology, namely the way it feels to see oneself as getting what one genuinely desires. The features of enjoyment that are naturally explained

by each of the leading theories of enjoyment are thus all naturally explained by reliabilist hedonism, and the enjoyment problem is dissolved.

I conclude, therefore, that reliabilist hedonism constitutes a revival of hedonism. It provides the hedonist with plausible solutions to at least three significant problems that are often taken by her opponents as providing sufficient reason to reject her view. Thus, if these opponents still wish to reject hedonism, it seems they must do so now by way of a different route.

9 Do Fitting Emotions Tell Us Anything About Well-Being?

Having introduced and defended reliabilist hedonism, I want to turn attention now to a rather different debate relating to the nature of well-being. This is the debate over the nature of the connection between emotions and well-being. As we will see, the possibility of reliabilist hedonism helps to show that what is seen by some in this debate to constitute a decisive refutation of hedonism is no refutation of the theory at all.

Recall, once more, that theories of well-being or welfare essentially aim to tell us which things are basically good and bad for us—or which things are the basic constituents of our well-being and ill-being. According to some theories, all and only experiences of a certain kind can directly contribute to our well-being. No doubt the most popular such theory is *hedonism*, which, recall, claims that all and only enjoyable experiences directly contribute to a subject's well-being, and all and only unpleasant experiences directly contribute to a subject's ill-being.[99]

While experiential theories like hedonism have been under attack for some time, this attack has recently become more pronounced.[100] The problem for these theories is that they seem to predict that nothing external to a subject's experiences should have any direct influence on the subject's well-being. After all, if well-being consists in (e.g.) enjoyable experience, then facts about the person's life that do not directly impact their enjoyable or

[99] For contemporary defenses of hedonism, see Crisp 2006 (ch. 4) and Feldman 2004 (ch. 4). An alternative experiential theory, "experientialism," has also recently been defended by Richard Kraut (2018).

[100] See e.g. Plato 1993, G. E. Moore 1903 (ch. 3), and W. D. Ross 1930 (ch. 5).

unpleasant experiences should be irrelevant to their well-being. Recently, however, it has been argued that facts of this kind can indeed directly impact a subject's well-being.[101] One such argument has come from Tobias Fuchs, who appeals to the following case:

> Imagine two women, both under the impression of being in a loving marriage. Their lives are identical apart from the fact that one of them [call her Jess] has the right impression, while the other one [Ingrid] is mistaken — she is being lied to and cheated on by her husband. (2018, p. 139)

Theories like hedonism appear to imply that Jess and Ingrid should seem to us to be equally well off. That's because hedonism claims that well-being consists in a certain kind of experience, and Jess and Ingrid each have the same experiences. The theory therefore implies that the two have equal levels of well-being; and so, the hope for the hedonist will no doubt be that our intuitions accord with this implication.

The problem, of course, is that our intuitions about the case are not as hedonism seems to predict: Jess intuitively seems better off than Ingrid. Since Jess seems better off than Ingrid merely in virtue of some fact about her life that does not directly impact her enjoyable experiences, the veracity of our intuition here seems to conflict with the truth of hedonism. And so, either our intuition or hedonism must be given up. While this argument is simple enough, it is also quite powerful. It leaves the hedonist in the uncomfortable position of having to "bite the bullet," either by suggesting that the intuition is for some

[101] Most recently, see e.g. Fuchs 2018, Lin 2014, Lin 2016, and Stevenson 2018. These arguments are of course inspired by Nozick 1974 (pp. 42-45).

reason misleading, or claiming that they themselves simply do not have the same intuition as many of us.

Regardless of how plausible we take these potential responses of the hedonist to be, Fuchs has offered what he takes to be a principled argument against them. Indeed, Fuchs offers a "working test" for well-being, which aims to provide an independent standard for testing when exactly a subject's well-being has been impacted. The idea behind the test is that, when we feel compassion for someone in virtue of some fact about them, and our compassion also meets a certain standard of correctness—when it is also the "right" sort of emotion to feel—the subject's well-being is negatively impacted in virtue of that fact about them. So since it seems right, for example, to feel compassion for Ingrid in virtue of her being deceived, the test apparently tells us that Ingrid is worse off in virtue of her being deceived. Given it reaches this conclusion by appealing only to facts about the correctness of compassion, and not to any claims or intuitions about well-being, Fuchs's test seems to provide an independent and conclusive way of both testing for changes in well-being, and refuting theories like hedonism.

In this chapter, I'll argue that Fuchs's test does neither of these things. In the following section, I'll offer a more detailed explanation of Fuchs's test and the relevant standard of correctness it appeals to. I'll then explain where Fuchs's application of the test goes wrong, and argue that, in virtue of a certain epistemological limitation, the test can offer us little help in the debate over the nature of well-being.

9.1 Fuchs's test and its significance

The general idea behind Fuchs's test, I have said, is that if compassion meets some standard of correctness—if it is the "right" sort of emotion to feel—then we are free to conclude that

the relevant subject is in some respect badly off. But what exactly is the relevant standard of correctness here that our compassion must meet?

The standard of correctness here is the emotion's *fittingness*. Emotions are typically taken to have not only affective and motivational components, but evaluative components as well. To see this, consider fear. Fear is certainly marked by a certain affect, and also disposes those who feel it to perform certain actions. Yet fear also has an evaluative component: it involves apprehending the *target* of that fear, or the thing at which the fear is directed, as instantiating a certain evaluative property—a property that is sometimes called the emotion's *formal object*.[102,103] In the case of fear, the formal object is the property of *being dangerous* (or being a danger to the fearer). This is evidenced by the fact that it would not make sense for me to feel fear if I did not take the target of that fear to be dangerous: if I did not experience the target as dangerous, there would be nothing to make my experience of fear intelligible. We can then say that my emotion "makes sense" just in case I see the target as instantiating the emotion's formal object. Importantly, however, a subject's emotion can sometimes make sense, even though we do not think the emotion is correct or fitting. If, for example, I know that a nearby spider is harmless, yet I fear it just the same, we should say that although my emotion makes sense, because I do see the spider as being dangerous, it is not correct or fitting, because the spider is not *in fact* a danger to me. If, in contrast, I am fearful of a nearby wild tiger, which I know is likely to attack me, my fear not only makes sense, because I see the tiger as being dangerous, but is correct or fitting, because the tiger is in fact a danger to

[102] See Deonna and Teroni 2012 (p. 41) and Fuchs 2018 (p. 135).

[103] I use "evaluative" here, rather than (say) "cognitive" or "representational," in order to remain as neutral as possible as to which particular competing theory of the nature of emotion is correct.

me. More generally: an emotion is fitting if and only if the target of the emotion in fact has the property that serves as the emotion's formal object.[104]

After outlining this notion of fittingness, Fuchs applies it to compassion. He notes that it does not make sense to feel compassion for someone whose life we take to be in all respects going well. If, for instance, your friend was by all accounts happy, healthy, wealthy, and wise, you may think it odd of me to report feeling compassion for her. Yet if I told you that she will soon be diagnosed with a serious illness, my compassion will seem perfectly in order. While this alone may not show that the property of *being worse off* is compassion's formal object, it at least suggests that a reduction in well-being is a necessary condition for compassion to be fitting. That is, it suggests that compassion's formal object, whatever it may be, must overlap with the property of *being worse off* at least to the extent that, whenever it is instantiated, the property of *being worse off* is instantiated as well. Together with the outlined notion of fittingness, this yields Fuchs's test for well-being:

> If compassion for some person A is fitting in virtue of A's having (or lacking) some property P, then A's well-being is negatively affected by A's having (or lacking) that property P.[105]

Since an emotion is fitting if and only if its target in fact has the property that serves as its formal object, it follows, so long as the fittingness of compassion requires that the target has the property of *being worse off*, that if compassion is fitting in virtue of the target's having or

---

[104] See D'Arms and Jacobson 2000 (pp. 65-90) and Deonna and Teroni 2012.

[105] Fuchs 2018 (p. 137).

lacking some property, then the target of the compassion is worse off in virtue of their having or lacking that property. The fittingness of compassion is then a sufficient condition for welfare having been diminished.

Having established his test, Fuchs suggests a way of utilizing it. He writes:

> We compare two lives, similar in all respects but that one has a property the other one lacks. Now we see whether compassion is fitting towards one of them, in virtue of having (or lacking) that property. If it is, we know that having (or lacking) that property is negatively affecting that person's well-being. (2018, p. 137)

The test can then be used to support our intuition that Ingrid is worse off than Jess. Since Ingrid is being cheated on by her husband and Jess is not, it follows, so long as Ingrid is the fitting target of compassion in virtue of being cheated on by her husband, that Ingrid is worse off than Jess in virtue of being cheated on by her husband. Consequently, it seems, the fact that a subject is deceived can negatively impact the subject's well-being.

As I've mentioned, Fuchs takes his test to provide us with an independent standard for changes in well-being whose authority cannot simply be dismissed by those whose theoretical commitments clash with its implications. That's because fittingness tracks the truth: an emotion's formal object must in fact be instantiated in order for the emotion to be fitting. Accordingly, so long as we establish that compassion would be fitting in some case, and that compassion's fittingness requires that its target is in some regard badly off, the claim that the target of the compassion is nevertheless no worse off is not a response that is available to the would-be objector. After all, by accepting that the compassion is fitting, we

accept that the target is worse off. The verdicts of the test then cannot plausibly be refuted on the grounds of conflicting theoretical commitments.

Now, that may be right, but notice how Fuchs actually applies his test:

> It seems that compassion is fitting towards [Ingrid], in virtue of her spouse's cheating on her and deceiving her. If we overheard a conversation in a diner of someone telling his friend about his sister being in Ingrid's situation, and he expressed compassion for his sister, surely we would not find this unintelligible — the stranger's compassion is perfectly fitting here. According to our compassion test, it follows that the woman is badly off in virtue of her being deceived and cheated on by her spouse. (2018, p. 139)

Here, Fuchs appeals to the fact that compassion for Ingrid seems fitting, and concludes on that basis that Ingrid is worse off. Notice, however, that the test says nothing about whether compassion *seems* fitting; only something about whether compassion *is* fitting. To be sure, many of us will share Fuchs's intuition that compassion seems fitting, but this intuition does not tell us that compassion is in fact fitting. It tells us only that compassion *seems* fitting. And, of course, it does not follow from the fact that compassion *seems* fitting that compassion *is* fitting—given it does not follow from the fact that a subject seems to have some property that the subject has that property—and so it does not follow from Fuchs's arguments that Ingrid's well-being is diminished.

The task for Fuchs, then, is to show that compassion for Ingrid *is* fitting. The problem here, however, is that we simply cannot know whether compassion for a subject is fitting without knowing whether the subject's well-being is diminished. To see why, return to

the example of fear. In the case of the nearby tiger, we can conclude that my fear is fitting, because we know that the tiger is a threat to me. And, in the case of the nearby spider, we can conclude that my fear is not fitting, because we know that the spider is not a threat to me. But now consider a case where we're not sure whether the animal is a threat to me. Suppose there's a nearby dog, which gives me some reason to perceive it as a threat—say, that it appears to be of an "aggressive" breed—and which I therefore fear, despite being unsure of how likely it is to attack. Since I see the dog as a threat, my emotion here surely makes sense. But is my emotion here also fitting? Well, of course, it depends. In particular, it depends on whether the dog is in fact a threat to me—on whether it is sufficiently likely to attack and harm me. If we do not know that it is in fact a threat to me, then we cannot conclude that my fear is fitting. The most we can conclude is that, given the dog appears to me to be of an aggressive breed, there is at least one reason for thinking that my fear may be fitting.

In the same way, I think, the most we can conclude about Fuchs's case is that, given Ingrid appears to us to be worse off, there is at least one reason for thinking that compassion for her may be fitting. After all, whether compassion is fitting depends on whether the target's welfare is diminished; and so, in order to *know* whether compassion for someone is fitting, we must *know* whether their welfare is diminished. Yet how could we know whether someone's welfare is diminished, without the true theory of welfare? The answer, outside of any case where it is already indisputable that welfare is or isn't diminished, is that we cannot. If we do not already know whether a subject's well-being is diminished in some case—if it is not a case, such as one involving sustained and brutal torture, where it could not plausibly be denied that the subject's welfare is diminished—then we cannot presently know whether compassion for the subject is fitting, because we do not know

whether the subject's well-being is diminished. This means that, so long as we do not know what well-being consists in, any case where we do not already know whether well-being is diminished will also be a case where we do not know whether compassion is fitting. And so, since we do not know what well-being consists in, and since we do not already know, in Fuchs's case, whether Ingrid's well-being is diminished, it follows that we do not know whether compassion for her is fitting. We then cannot conclude that compassion for her is fitting, and so cannot employ Fuchs's test here. The most we can conclude is that, given Ingrid seems to us to be worse off, there is at least one reason for thinking that compassion for her may be fitting. Just as in the case of my fear of the dog, we cannot know whether our emotion is fitting without knowing whether the target has the relevant property.

(It is worth briefly mentioning that this epistemological limitation will apply, for analogous reasons, to any emotion whose fittingness requires some change in its target's well-being. So long as we do not know either what well-being consists in, or whether the target's well-being has been changed in the relevant way, we cannot know whether the emotion is fitting. I will return to this point below.)

Importantly, in making this argument, I do not mean to appeal to any kind of extreme skepticism about what we can or do know about well-being. As I've suggested, I think there are at least some cases, such as those involving brutal torture, where we do know that a subject's well-being is diminished, despite our not knowing what well-being consists in. As it happens, I think these are also just the cases where we can surely utilize Fuchs's test: since we know that welfare is diminished, we can conclude that compassion is fitting, and so can conclude via Fuchs's test that welfare is diminished. As this exercise shows, however, the cases where we can surely utilize his test are just those where the test is needed least. Since we cannot know whether compassion is fitting without knowing whether welfare is

diminished, we cannot confidently apply Fuchs's test without already knowing the very conclusion that the test is meant to help us reach. Seen in this way, Fuchs's test and its implications for the debate over the nature of well-being are likely to seem far less significant than Fuchs sometimes suggests them to be. At the very least, it seems clear that the test does not prove to us that experiential theories like hedonism are false, since the cases that would allow the test to prove this are just those where we do not know whether compassion is fitting.

But, even if that is right, it may be objected that the test still provides *some* independent reason for rejecting these views. After all, in cases like Ingrid's, many of us will agree that compassion for Ingrid seems fitting. And, it may be argued, while this may not *prove* the claim that compassion for Ingrid is fitting, it at least *justifies* that claim—it allows us to justifiably believe it, given our emotion does not seem subject to any defeaters, such as hallucination or drunkenness. In that case, it seems we are also justified in claiming that Ingrid's welfare is diminished, since compassion is fitting only if a subject's welfare is diminished. The objection, then, is that even if Fuchs's test faces the epistemological limitation I discussed, the fact that compassion for Ingrid seems fitting still provides at least some reason for thinking that she is worse off, and so at least some reason for thinking that theories like hedonism are false.

I'm willing to concede that compassion for Ingrid seems fitting, and compassion for Jess does not. I'm also willing to concede that this provides some reason for thinking that Ingrid is (or, at least, will be) worse off than Jess. What I am not willing to concede, though, is that any of this provides a reason for thinking that hedonism is false. In fact, I think it is more likely to provide a reason for thinking that we have misidentified the property in virtue of which compassion for Ingrid seems fitting. Since Ingrid and Jess are in loving and happy

marriages, being cheated on would almost certainly lead either of them to many, very unpleasant experiences. But because only Ingrid is being cheated on, only Ingrid is overwhelmingly likely to experience this great deal of unpleasantness in the future. As I see it, it seems far more plausible that it is in virtue of this fact that compassion for Ingrid seems fitting, and compassion for Jess does not, than it seems plausible that being deceived or cheated on would explain this difference.

Perhaps you disagree. Then consider a slightly modified version of the case. Suppose Ingrid and Jess's marriages are instead not happy, and that if either of them were to learn of their partner's infidelity, it would not lead to any unpleasantness at all. In fact, it would make them happy: they would see it as a relief, and would eventually look back on it positively. Would compassion for Ingrid, whose partner is in fact cheating on her, still seem fitting in this case? As I see it, while compassion might seem fitting in virtue of her being in an unhappy marriage (given the likely attendant unpleasantness), it would not seem fitting in virtue of her being deceived or cheated on. Since she would see it as a relief upon finding out about the deception, it would not cause her to experience any immediate unpleasantness; and since she would look back on it positively, it would not cause her to experience any unpleasantness in the long term. Even if you disagree that compassion does not seem fitting in virtue of Ingrid's being deceived or cheated on here, though, I'm confident you'll agree that it seems far *less* fitting in this case than it does when she is in a happy marriage. Yet, in each case, she is equally cheated on and deceived; and so, if these are indeed ultimately the properties in virtue of which compassion for her is fitting in one case, compassion for her should seem equally fitting in the other case. That is, if the fact that ultimately explains why compassion for Ingrid is fitting—and thus why she is worse off—when her marriage is happy is that she has the property of *being deceived*, then compassion for Ingrid should seem

equally fitting when her marriage is not happy, because that fact applies to her just the same. Since compassion for Ingrid does not seem equally fitting in the cases, it seems considerably less plausible that this is ultimately the property in virtue of which compassion for her is fitting in either case. And if this is not the property in virtue of which compassion for Ingrid is fitting, then, while the fact that compassion for Ingrid seems fitting and compassion for Jess does not may provide some reason for thinking that Ingrid is (or, at least, will be) worse off than Jess, that fact does not also provide a reason for thinking that deception, or something outside of Ingrid and Jess's experiences, is what ultimately explains this difference in well-being. Contrary to the objector's suggestion, then, our intuitions about compassion's fittingness do not provide a reason for thinking experiential theories like hedonism are false. In fact, given that the case where compassion for Ingrid seems more fitting also happens to be the one where she seems more likely to experience more unpleasantness, and the case where compassion for her seems less fitting also happens to be the one where she seems less likely to experience any unpleasantness, the hedonist (and experientialist more generally) seems particularly well-positioned to plausibly explain why compassion seems more or less fitting in the cases.[106]

As my arguments here illustrate, our intuitions about compassion's fittingness are in no better position to tell us either when changes in well-being have occurred, or which properties ultimately explain these changes, than our everyday intuitions about changes in

[106] One such explanation is provided by reliabilist hedonism, or the view presented and defended in chapters 7 and 8. That is, the hedonist can explain our intuitions about these cases by appeal to the claim that our evaluative intuitions track which things have proven themselves to us to reliably lead to enjoyment. Compassion seems more fitting in the case where Ingrid's life has more features that have proven themselves to us to be reliable indicators of future unpleasantness, and it seems less fitting in the case where her life has fewer features that have proven themselves to us to be reliable indicators of future unpleasantness, because our intuitions about compassion's fittingness track our intuitions about changes in well-being, and our intuitions about changes in well-being track which things have proven themselves to us to reliably lead to enjoyable or unpleasant experience.

well-being. While these intuitions about compassion's fittingness may reinforce what our intuitions about changes in well-being can already tell us, we should be careful not to overstate their significance. They surely do not provide conclusive evidence of anything that would significantly advance the debate over the nature of well-being (such as the falsity of experiential theories), and indeed are just as likely to mislead us with respect to determining which things are basically good and bad for us as our everyday intuitions about changes in well-being. Accordingly, if our aim in this debate is to determine which things are basically good and bad for us, appealing to the fittingness of compassion does not seem an especially promising way forward.

Really, these points apply to any putative "test" for well-being that appeals to the fittingness of an emotion whose fittingness requires some change in its target's welfare. That's because, since fittingness tracks the truth, we will be unable to conclude that any such emotion is fitting without knowing whether the target's well-being has been impacted in the relevant way, leaving us to retreat to our intuitions about whether the target's well-being has been so impacted, and thus about whether the emotion is fitting, in any case where these things are not already known. While our intuitions here may provide some evidence for claiming that the target's well-being has been impacted, they may also provide misleading evidence about the ultimate explanation for that impact. Since the debate over the nature of well-being is essentially concerned with determining which things are basically good and bad for us—or determining which things ultimately explain changes in well-being—these "tests" therefore seem no better positioned to significantly advance the debate than appeals to our everyday intuitions about changes in well-being. I hope to have shown here that, while it may at times be tempting to think otherwise, significant progress in this debate still requires

the difficult work of reasoning our way, via our intuitions, to the true theory of well-being. Fitting emotions can provide no shortcut.

REFERENCES

Adams, R. M. (1999). *Finite and Infinite Goods*. Oxford: Oxford University Press.

Aristotle. (2000). *Nicomachean Ethics*. Roger Crisp (trans.). Cambridge: Cambridge University Press.

Arpaly, N. and T. Schroeder. (2014). *In Praise of Desire*. New York: Oxford University Press.

Badhwar, N. K. (2014). *Well-Being: Happiness in a Worthwhile Life*. New York: Oxford University Press.

Berridge, K., and M. Kringelbach (2008). "Affective Neuroscience of Pleasure: Reward in Humans and Animals." *Psychopharmacology* 199: 457-480.

Bramble, B. (2013). "The Distinctive Feeling Theory of Pleasure." *Philosophical Studies* 162: 201-217.

Brandt, R. (1979). *A Theory of the Good and the Right*. Oxford: Clarendon Press.

Chang, R. (2016). "Parity, Imprecise Comparability and the Repugnant Conclusion." *Theoria* 82: 182-214.

Craig, E. (1990). *Knowledge and the State of Nature*. Oxford: Clarendon Press.

Crisp, R. (2006). *Reasons and the Good*. Oxford: Oxford University Press.

D'Arms, J., and D. Jacobson (2000). "The Moralistic Fallacy: On the 'Appropriateness' of the Emotions." *Philosophy and Phenomenological Research* 61: 65-90.

Darwall, S. (2002). *Welfare and Rational Care*. Princeton: Princeton University Press.

Davis, W. (1981). "A Theory of Happiness." *American Philosophical Quarterly* 18: 111-120.

De Brigard, F. (2010). "If You Like It, Does It Matter If It's Real?" *Philosophical Psychology* 23: 43-57.

Deonna, J. A., and F. Teroni (2012). *The Emotions: A Philosophical Introduction*. New York: Routledge.

Dogramaci, S. (2012). "Reverse Engineering Epistemic Evaluations." *Philosophy and Phenomenological Research* 84: 513-530.

---. (2015). "Communist Conventions for Deductive Reasoning." *Noûs* 49: 776-799.

Feldman, F. (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Clarendon Press.

---. (2010). *What Is This Thing Called Happiness?* Oxford: Oxford University Press.

---. (2011). "What We Learn from the Experience Machine." In R. M. Bader and J. Meadowcroft (eds.), *The Cambridge Companion to Nozick's 'Anarchy, State, and Utopia'*, pp. 59-88. Cambridge: Cambridge University Press.

Fletcher, G. (2013). "A Fresh Start for the Objective-List Theory of Well-Being." *Utilitas* 25: 206-220.

Fuchs, T. (2018). "A Working Test for Well-Being." *Utilitas* 30: 129-142.

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Hare, C. (2013). *The Limits of Kindness*. Oxford: Oxford University Press.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.

Haybron, D. M. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.

Heathwood, C. (2006). "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128: 539-563.

---. (2007). "The Reduction of Sensory Pleasure to Desire." *Philosophical Studies* 133: 23-44.

---. (2011). "Desire-Based Theories of Reasons, Pleasure, and Welfare." *Oxford Studies in Metaethics* 6: 79-106.

---. (2019). "Which Desires Are Relevant to Well-Being?" *Noûs* 53: 664-688.

Hewitt, S. (2010). "What Do Our Intuitions about the Experience Machine Really Tell Us about Hedonism?" *Philosophical Studies* 151: 331-349.

Holton, R. (2003). "How Is Strength of Will Possible?" In S. Stroud and C. Tappolet (eds.), *Weakness of Will and Practical Irrationality*, pp. 39-67. Oxford: Clarendon Press.

Hurka, T. (2011). *The Best Things in Life: A Guide to What Really Matters*. Oxford: Oxford University Press.

Kagan, S. (1992). "The Limits of Well-Being." *Social Philosophy & Policy* 9: 169-189.

---. (1994). “Me and My Life.” *Proceedings of the Aristotelian Society* 94: 309-324.

---. (2009). “Well-Being as Enjoying the Good.” *Philosophical Perspectives* 23: 253-272.

Kawall, J. (1999). “The Experience Machine and Mental State Theories of Well-Being.” *Journal of Value Inquiry* 33: 381-387.

Keller, S. (2009). “Welfare as Success.” *Noûs* 43: 656-683.

Kraut, R. (2007). *What Is Good and Why: The Ethics of Well-Being*. Cambridge: Harvard University Press.

---. (2018). *The Quality of Life: Aristotle Revised*. Oxford: Oxford University Press.

Lewis, D. (1988). “Desire as Belief.” *Mind* 97: 323-332.

Lin, E. (2014). “Pluralism about Well-Being.” *Philosophical Perspectives* 28: 127-154.

---. (2016). “How to Use the Experience Machine.” *Utilitas* 28: 314-332.

---. (2017a). “Against Welfare Subjectivism.” *Noûs* 51: 354-377.

---. (2017b). “Enumeration and Explanation in Theories of Welfare.” *Analysis* 77: 65-73.

---. (2018). “Attitudinal and Phenomenological Theories of Pleasure.” *Philosophy and Phenomenological Research*. DOI 10.1111/phpr.12558.

Mill, J. S. (1863). *Utilitarianism*.

Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Nagel, T. (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

---. (2011). *On What Matters. Vol. 1*. Oxford: Oxford University Press.

---. (2016). “Can We Avoid the Repugnant Conclusion?” *Theoria* 82: 110-127.

Plato (1993). *Philebus*. Dorothea Frede (trans.). Indianapolis: Hackett Publishing.

Railton, P. (1986). “Facts and Values.” *Philosophical Topics* 14: 5-31.

---. (1989). “Naturalism and Prescriptivity.” *Social Philosophy and Policy* 7: 151-174.

Reynolds, S. (2017). *Knowledge as Acceptable Testimony*. Cambridge: Cambridge University Press.

Ross, W. D. (1930). *The Right and the Good*. Oxford: Clarendon Press.

Schapiro, T. (2009). "The Nature of Inclination." *Ethics* 119: 229-256.

Schroeder, M. (2010). "Value and the Right Kind of Reason." *Oxford Studies in Metaethics* 5: 25-55.

Schroeder, T. (2004). *Three Faces of Desire*. New York: Oxford University Press.

Setiya, K. (2012). *Knowing Right from Wrong*. Oxford: Oxford University Press.

Sidgwick, H. (1907). *The Methods of Ethics*. London: Macmillan.

Silverstein, M. (2000). "In Defense of Happiness: A Response to the Experience Machine." *Social Theory and Practice* 26: 279-300.

Sinhababu, N. (2009). "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118: 465-500.

---. (2017). *Humean Nature*. Oxford: Clarendon Press.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stevenson, C. M. (2018). "Experience Machines, Conflicting Intuitions and the Bipartite Characterization of Well-Being." *Utilitas* 30: 383-398.

Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109-166.

Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.

Velleman, J. D. (1991). "Well-Being and Time." *Pacific Philosophical Quarterly* 72: 48-77.

Weijers, D. (2014). "Nozick's Experience Machine Is Dead, Long Live the Experience Machine!" *Philosophical Psychology* 27: 513-535.

Weijers, D. and V. Schouten (2013). "An Assessment of Recent Responses to the Experience Machine Objection to Hedonism." *Journal of Value Inquiry* 47: 461-482.

Woodard, C. (2013). "Classifying Theories of Welfare." *Philosophical Studies* 165: 787-803.

---. (2015). "Hybrid Theories." In G. Fletcher (ed.), *The Routledge Handbook of the Philosophy of Well-Being*, pp. 161-174. Routledge.