

Happiness and Welfare

by

Sean Michael Meseroll

Submitted to the graduate degree program in Philosophy
and the Graduate Faculty of the University of Kansas in partial fulfillment of the
requirements for the degree of Doctor of Philosophy.

Chairperson Dale Dorsey

Ben Eggleston

Erin Frykholm

John Symons

Nancy Hamilton

Date Defended:

May 14, 2014

The Dissertation Committee for Sean Michael Meseroll
certifies that this is the approved version of the following dissertation:

Happiness and Welfare

Chairperson Dale Dorsey

Date Approved:

May 14, 2014

Abstract

In this dissertation I argue that while hedonism seems to be the correct theory of happiness, happiness does not seem to be the essence of welfare; after all, it appears that a person may be brainwashed over a given duration, may be happy over that same duration, but not also be well off over that duration, all things considered; this suggests that well-being consists of capacity-fulfillment.

Hedonism about happiness (HH), maintains that you are happy to the extent that you have pleasure, unhappy to the extent that you have pain. Besides hedonism about happiness, there are three popular theories of happiness: desire-satisfaction, life-satisfaction, and emotional state. I judge these theories both by whether they accord with our commonsense intuitions and whether they have any internal problems. All three conflict with some of our intuitions and have internal problems. HH, however, accords with our intuitions and is not susceptible to the internal problems that the other theories are susceptible to.

The relationship between happiness and well-being is a complex one. Happiness does seem intrinsically good for the happy: whenever you experience an episode of happiness, you seem thereby better off. By appealing to the case of brainwashing, however, I show that happiness seems neither necessary nor sufficient for welfare.

Before introducing my own theory of welfare, I discuss five alternatives: desire-satisfaction, life-satisfaction, self-fulfillment, perfectionism, and objective list. The first four fail to accord with intuition in the case of brainwashing and the last three have internal problems. My own theory of welfare, capacity-fulfillment about welfare (CFW), maintains that you are well off to the extent that you successfully exercise your basic capacities. We possess four basic capacities: affection, cognition, conation, and locomotion. While CFW does not accord with

some of our intuitions, it accommodates the case of brainwashing and incorporates some of the strengths of the welfare theories I criticize while remaining immune to the internal problems they have.

Acknowledgements

I am greatly indebted to my family. Without their support and intervention, I would not be pursuing philosophy. This dissertation is primarily dedicated to them. I would also like to thank Dale Dorsey, my dissertation adviser. His guidance and encouragement have been invaluable. With his constructive criticism and astute feedback I was able to greatly improve the quality of my work. I want to thank my other committee members, too: Ben Eggleston, Erin Frykholm, John Symons, and Nancy Hamilton. I am honored to have all of you serve on my committee. I further owe thanks to the other faculty of the Philosophy Department at the University of Kansas: I first learned to do philosophy while participating in their seminars and being their teaching assistant. Finally, I would like to thank my fellow graduate students in the Philosophy Department: I received constructive feedback from conversations and colloquia. I owe thanks in particular to Micah Bailey: our conversations helped me gain greater conceptual clarity.

Table of Contents

Introduction.....	1
Chapter 1: Happiness Preliminaries.....	7
1.1 Preliminaries	7
1.1.1 Apparent and actual happiness	7
1.1.2 Descriptive versus evaluative senses of ‘happy’	9
1.1.3 Being occurrently, dispositionally, and predominately happy.....	10
1.1.4 Relational and non-relational happiness.....	12
1.1.5 Philosophical analysis of happiness.....	15
1.2 Examples of happiness	18
1.2.1 Brainwashed Sally	18
1.2.2 Mary the mother	19
1.2.3 Glenda the graduate student	19
1.2.4 Seneca the Stoic.....	20
1.2.5 Andrew the swine	20
Chapter 2: What Happiness is Not.....	22
2.1 Desire-satisfaction about happiness (DSH)	22
2.1.1 The essence of DSH	22
2.1.2 DSH and our five cases.....	25
2.1.3 Additional considerations.....	29
2.2 Life-satisfaction about happiness (LSH).....	32
2.2.1 The essence of LSH.....	32
2.2.2 LSH and our five cases	34
2.2.3 Additional considerations.....	39
2.3 The emotional state theory of happiness (EH).....	41
2.3.1 The essence of EH	41
2.3.2 EH and our five cases	43
2.3.3 Additional considerations.....	48
2.4 Chapter conclusion	57
Chapter 3: What Happiness Is	58
3.1 Hedonism about happiness (HH).....	58
3.1.1 The essence of HH.....	58
3.1.1.1 The essence of pleasure.....	63
3.1.2 HH and our five cases.....	66
3.1.3 Feeling and being happy	70
3.1.4 Defending HH against Haybron.....	71
3.2 Chapter conclusion	77
Chapter 4: Happiness and Welfare.....	78
4.1 Introduction to welfare.....	78
4.2 The benefit of happiness.....	82
4.3 Brainwashing.....	84
4.3.1 The essence of brainwashing	84
4.3.2 Brainwashed Sally	87
4.3.3 The harm of brainwashing	90
4.4 Hedonism about happiness and welfare (HHW).....	91
4.5 Authentic happiness theories of welfare.....	95

4.5.1 Authentic hedonism about happiness and welfare	96
4.5.1.1 Truth-adjusted HHW	96
4.5.1.2 Desert-adjusted HHW.....	99
4.6 Chapter conclusion	101
Chapter 5: What Welfare is Not.....	102
5.1 Desire-satisfaction about welfare (DSW).....	102
5.1.1 The essence of DSW	102
5.1.2 The pro tanto benefit of satisfied desires	103
5.1.3 DSW and Sally.....	103
5.1.4 Full-information DSW	105
5.1.5 Criticizing full-information DSW	107
5.2 Life-satisfaction about welfare (LSW).....	110
5.2.1 The essence of LSW	110
5.2.2 LSW and Sally	110
5.2.3 Authentic LSW	111
5.2.4 Additional criticisms of LSW	115
5.3 Self-fulfillment theory of welfare (SFW).....	117
5.3.1 The essence of SFW	117
5.3.2 SFW and Sally	120
5.3.3 SFW's responses to Sally.....	121
5.3.4 Additional criticisms of SFW	122
5.4 Perfectionism about welfare (PW).....	123
5.4.1 The essence of PW	123
5.4.2 Aristotle's PW.....	124
5.4.3 Hurka's PW.....	127
5.5 The objective list (OL).....	133
5.5.1 The essence of OL	133
5.5.2 OL and Sally.....	133
5.5.3 Support for OL.....	135
5.5.4 Criticisms against OL	136
Chapter 6: What Welfare Is	140
6.1 Capacity-fulfillment theory of welfare (CFW)	140
6.1.1 The essence of CFW	140
6.1.2 Basic capacities.....	142
6.1.3 CFW and brainwashed Sally	150
6.1.4 Strengths of CFW	153
6.1.5 Defending CFW from criticism.....	156
Conclusion	161
Bibliography	167

Introduction

Let us imagine a person, Sally, who over several months is thoroughly *pleased*. Never has she *felt* so spiritually uplifted, so alive with power, so at ease with people, so confident of the truth of her convictions, and just so *fulfilled*. Not only does she feel great: Sally also takes pleasure in feeling so great. Indeed, she feels so pleasant and takes so much pleasure in feeling so pleasant that Sally has a higher surplus of pleasure than displeasure, the highest surplus she has ever had over a similar duration. There is a hitch, however: Sally is a member of a destructive cult.¹ Although she is satisfied and feels so good, Sally is being brainwashed.

Now, imagine that Sally is fortunate enough to have a family that decides to intercede on her behalf, hiring professionals to conduct an intervention. Suppose that the intervention goes well, so well that Sally leaves the cult. Over the next month or two Sally struggles to cope with the loss of leaving the cult. She is often *displeased*: she *feels* spiritually deflated, alienated from others, unsure of what convictions to hold, let alone be confident of, and just so *unfulfilled*. She realizes that she was manipulated by the cult and hence feels anxiety over not knowing whom she can trust. This leaves Sally feeling downright miserable, so miserable that she has occasionally contemplates suicide.

Sally begins seeing a therapist to help ease her struggles. With the therapist's help she is now beginning to do something she could not do while brainwashed: exercise her own critical

¹ By 'cult', I merely mean "a deviant [...] organization with novel beliefs and practices. Deviance is departure from the norms of a culture in such a way as to incur the imposition of extraordinary costs from those who maintain the culture" (Rodney Stark and William Sims Bainbridge, *A Theory of Religion* [New Brunswick, NJ: Rutgers University Press, 1996], 124). I removed 'religious' from Stark and Brainbridge's definition; cults need not be religious, after all. There are all sorts of cults: business, commercial, political, and so on. Thus construed, 'cult' is a descriptive, not necessarily an evaluative, term: one may be part of a cult and not thereby be worse off. The same is *not* true of 'destructive cult', since it is evaluative: if one is a member of a destructive cult, one is thereby made worse off. This is because I use 'destructive cult' to refer to a cult that employs brainwashing on at least some of its members. Once we examine the nature of brainwashing in chapter two it should be clear why brainwashing is harmful to the brainwashed and thus is an evaluative term.

capacities, reflect over the life she wants to lead, consider what evidence she has for her beliefs, and so on. Exercising her capacities in this manner feels good to Sally, and she takes pleasure from once again relying on her own thinking instead of relying on another's.

Fast-forward several years: Sally is still out of the cult. And she no longer has difficulty coping, at least not nearly to the extent she once did. More often than not Sally is *pleased*: she *feels* content, sure of her convictions, and just *fulfilled*. Not only does she feel good, her feeling so good brings Sally pleasure. She also continues to exercise her critical capacities; indeed, she is now a better critical thinker than she was prior to her brainwashing. This state of affairs, too, brings her pleasure. Sure, there are things she wishes that she had done otherwise; but for the most part, she is content: if given the choice to continue the brainwashed-free life she now leads or go back to the brainwashed life, Sally would choose to continue her current life. That said, she did feel better and was more satisfied while brainwashed than she is now. The amount, intensity, and duration of the pleasure she experienced while brainwashing far surpasses the pleasure she experiences several years after leaving the cult behind. Now, during her brainwashing Sally did also experience some intense suffering: she lost some close friendships and distanced herself from her family. Indeed, she suffered more as a cult member than she suffers several years after leaving it. Sally just experienced so much more pleasure while brainwashed that the amount of pleasure minus suffering she had while brainwashed outweighs the net amount of pleasure she has years afterward.

Let us refer to the several months that Sally was being brainwashed as case 'A', refer to the first couple of months after Sally leaves the brainwashing as case 'B', and refer to the several years after Sally leaves the brainwashing as case 'C'. It seems to me that we are justified in making at least three intuitive claims about A, B, and C. For one, in A, even though Sally is

brainwashed, she seems *happy*, so happy that she seems happier in A than in B or C. For another, Sally's happiness *does* seem to make her pro tanto better off: she seems better off brainwashed and happy, other things equal, than brainwashed without as much happiness. Lastly, while Sally seems better off in A than in B, she seems better off in C than in A. That is to say, even though Sally's brainwashing seems to make her pro tanto worse off, it does not seem to make all-things-considered worse off: she seems better off in A than in B. And even though Sally's brainwashed happiness seems to make her pro tanto better off, it does *not* seem to make her all-things-considered better off: once we consider what the brainwashing does to Sally, she seems better off in C, years after she left the brainwashing.

I appeal to this example throughout my dissertation. I contend that the correct theories of happiness and welfare should accord with it. While it is not so difficult for theories of happiness to accord with it, we shall see that it is difficult for theories of welfare to accord with it.

This dissertation consists of six chapters. The first three concern the nature of happiness; the fourth concerns the relationship between happiness and welfare; and the last two chapters concern the nature of welfare. I begin the first chapter with a discussion of what I (and other scholars studying happiness) mean by the term 'happiness'. Once I discuss some other preliminaries I then offer what I intend to be are five uncontroversial commonsense examples of happiness—not just examples I find intuitive but examples I think *we* are prereflectively inclined to agree on non-inferential grounds are examples of happiness.

In chapter two I explicate three popular theories of happiness: desire-satisfaction, life-satisfaction, and psychic affirmation. I judge the theories both by whether they accord with the five examples and whether they have any internal problems. If I am correct, all three conflict with some of the cases. Indeed, life-satisfaction conflicts with intuition in three of the five cases,

making it a wildly implausible theory of happiness. While desire-satisfaction and psychic affirmation accord with intuition in four of the five cases, each has internal problems sufficient to make them implausible.

In chapter three I discuss a theory of happiness that accords with intuition in all five cases and is free of the problems that plague the other theories: hedonism about happiness (hereafter also referred to as ‘HH’). In a nutshell, it maintains that happiness is constituted by pleasure. HH is different from other hedonistic theories by remaining neutral about the essence of pleasure: rather than become engrossed in the debate over whether pleasure is a feeling or an attitude, HH just maintains whatever the essence of pleasure is, *that*, constitutes happiness. My thesis for chapter three, then, is: happiness consists only of pleasure—whatever pleasure turns out to be. Before concluding the chapter I defend HH against criticisms of the anti-hedonist Daniel Haybron, whose criticisms have yet to be fully addressed in the literature.

In chapter four I consider the relationship between happiness and well-being, focusing on both whether a person need be happy in order to be well off and whether happiness is all that is required to be well off. It turns out that the relationship between happiness and welfare is a complex one. Happiness does seem intrinsically good for the happy: whenever you experience an episode of happiness, you seem thereby better off. That said, happiness does *not* seem necessary for welfare in the sense that ϕ may benefit you without thereby increasing your happiness. Moreover, happiness does *not* seem sufficient for well-being: it seems that you may enjoy a high level of happiness without also enjoying a correspondingly high level of welfare. The case of Sally is a prime example. She seems better off years after she leaves the cult than she does while brainwashed even though she is less happy. My theory of brainwashing is a

modification of the sociologist Benjamin Zablocki's theory. If I am right, brainwashing is the set of transactions whereby one person transforms the other into her deployable agent.

I conclude chapter four with a discussion of *authentic* happiness. Someone who believes that welfare is happiness may appeal to the notion of authenticity, arguing that authentic happiness is the essence of welfare, or that it is at least more prudentially valuable than *inauthentic* happiness. So I modify HH to try and make it accord with the case of Sally, appealing to Fred Feldman's truth-adjusted and desert-adjusted hedonistic theories of welfare. Authentic hedonism about happiness, however, seems false because both truth-adjusted and desert-adjusted theories fail to accord with intuition in the case of Sally. Thus, happiness—authentic or not—seems neither necessary nor sufficient for welfare.

This brings us to chapter five, where I begin my discussion of welfare, as such. Here I continue my discussion of what welfare is *not*. I explicate and then criticize five theories of welfare: desire-satisfaction, life-satisfaction, self-fulfillment, perfectionism, and objective list. The first three fail to accord with intuition in the case of Sally. Once modified to distinguish between *authentic* and *inauthentic* happiness, however, desire-satisfaction and life-satisfaction accord with intuition in the case of Sally. But they are only able to accord with intuition by becoming more like my own theory of welfare. Moreover, even though they now accord with intuition, each has internal problems. And while perfectionism and objective list can be formulated to accord with intuition in the case of Sally, they only do so by becoming more like my own theory; plus, they suffer from problems my own theory of welfare is not susceptible to.

In my last chapter, chapter six, I introduce my own theory of welfare: capacity-fulfillment about welfare (hereafter also referred to as 'CFW'). If it is right, you are well off to the extent that you successfully exercise your basic capacities. As I see it, we possess four basic

capacities: affection, cognition, conation, and locomotion. All of our non-basic capacities can be explained by one or more of our basic capacities. CFW accords with intuition in the case of Sally and incorporates some of the strengths of the welfare theories I criticize while remaining immune to the internal problems they have. For example, CFW accords with the world as well as some objective lists; yet CFW is *not* merely a list of purported goods: CFW is also an explanation for why those goods are intrinsically good for us. I conclude the chapter by defending CFW against objections.

The thesis, then, that I defend in this dissertation is: hedonism seems to be the correct theory of happiness in that pleasure seems to be the essence of happiness; but happiness does not seem to be the essence of welfare, since it appears that a person may be brainwashed over a given duration, may be happy over that same duration, but not also be well off over that duration, all things considered; this suggests that well-being consists of capacity-fulfillment. The capacity theory I have in mind maintains that you are well off to the extent that you successfully exercise your basic capacities: affection, cognition, conation, and locomotion. With this said, let us begin.

Chapter 1: Happiness Preliminaries

1.1 Preliminaries

1.1.1 Apparent and actual happiness

We use the term ‘happy’ in several different ways.² Sometimes people say that they *feel* happy. Other times people say that they *are* happy. When talking about others, sometimes we even say that they *seem* happy. What is going on here?—that is, are there distinctions between being, feeling, and seeming to be happy, or are they all the same phenomenon? In this section I aim to show that while it is *not* clear whether there is a distinction between feeling happy and being happy, it is clear that there *is* a distinction between being happy and merely seeming to be happy, between *actual* happiness and *merely apparent* happiness.

Imagine Blake, a hard working accountant for a tax accounting firm. It so happens that it is tax season, Blake’s busiest time of the year. In order to please his customers (and his bosses), Blake makes sure to don both a smile and a cheerful demeanor. His customers walk away feeling satisfied over saving money and having a pleasant tax representative to work with. Unbeknownst to his customers, however, Blake is *not* feeling very pleasant: he is not pleased to work at his firm, especially during tax season. He also feels dissatisfied in general for not putting his graduate degree in history to good use. In this scenario, it would be strange to describe Blake as happy. While he may appear happy from his demeanor, this is merely a façade: he does it merely to stay employed. He merely seems happy; he is *not* actually happy.

Cases like this illustrate that there is a difference between the predicates ‘seems happy’ and ‘is happy’. The first denotes seeming to be happy, *apparent* happiness, while the latter

² In the following five sections of chapter one I do not focus on unhappiness. This is for brevity’s sake. For each discussion of happiness, the same can be made for unhappiness, unless otherwise noted. For instance, just as we can distinguish actual from merely apparent happiness, we can distinguish actual from merely apparent *un*happiness.

denotes being happy, *actual* happiness.³ Now, this of course is not to say that one cannot both seem to be happy and be happy at the same time. Indeed, most people who are happy seem happy, too. That is, most who are happy display signs that make it look as though they are happy. But the point is: not all people who seem happy are actually happy.⁴ There of course appears to be a common set of behavior we associate with happiness; but that set of behavior is *not* identical with happiness itself since one can portray the common signs of happiness without actually being happy. What I am interested in explicating is, *actual*, not merely apparent, happiness.

Unfortunately, while it is clear that one can seem happy without being happy, it is not so clear that one can feel happy without also being happy.⁵ Every time I have said that “I feel happy” I think I also meant that I am happy; but while it would seem odd for one to say that “I feel happy, but I am not happy”, the sentence does not appear to imply a contradiction. The term ‘I feel happy’ does not appear to imply that you are happy.⁶ How about the other way around: can you be happy without feeling happy? Does one contradict oneself if one says, “I am happy,

³ Davis (1981a: 111) makes this distinction, too.

⁴ I said previously that as there is a difference between actual and merely apparent happiness, there is also a difference between actual and merely apparent unhappiness. But while the prospect of someone feigning happiness even though she is not happy seems commonplace, the prospect of someone feigning unhappiness even though she is not unhappy may seem odd. I grant this oddity: it seems far less common that people feign unhappiness than they feign happiness. But the same conceptual distinction applies: it is possible for one to seem to be unhappy and yet not be unhappy. Hence, there is a distinction between actual and merely apparent unhappiness, too.

⁵ See Wayne Sumner; he discusses the phenomena of feeling happy, but does not discuss how we understand what the connection is between feeling happy and being happy (1996: 144-145).

⁶ Davis and Christine Vitrano apparently disagree. Davis implies that the relationship between feeling and happy and being happy is *not* that of feeling sick and being sick: while one may feel sick and actually not be sick, one may not feel happy unless one is also actually happy (Davis 1981a: 111). But he fails to offer an argument to support his claim. Vitrano maintains that just as one cannot feel scared or feel angry without being scared or being angry, one cannot feel happy without also being happy (Vitrano 2013: 112). Happiness, according to Vitrano, is like any other emotion: if you are feeling it, “there is no principled way to distinguish the “feeling” from the “being”” (Ibid.). For reasons discussed below (see section 2.3), I doubt, however, that happiness is an emotion.

but I do not feel happy”)? It is not obvious that one does. I am puzzled by this. The correct account of happiness should help solve this puzzle.

1.1.2 Descriptive versus evaluative senses of ‘happy’

It is all well and good to discuss merely apparent versus actual happiness; but we also need to discuss the difference between descriptive and evaluative senses of ‘happy’. We use ‘is happy’ in at least three different ways. Sometimes we use it to express a property that is *good for* the person who has it, where ‘Sally is happy’ expresses something like the proposition that Sally is leading a life that is good *for her*. Thus construed, ‘is happy’ is *evaluative*, since it implies that happiness is valuable for the person who has it. But this judgment does not necessarily express a fact about all value, taking into account moral, aesthetic, and perhaps other value; all it tells us is that Sally’s life is high in a particular type of value: *prudential* value. There is another evaluative sense of ‘is happy’ that does seem to express a fact about all value. Think of the sentence ‘Sally is leading a happy life’. It seems to express something like the proposition that Sally is leading an *overall* good life. The idea is that Sally does not merely live a life that is good for her, but an overall good life: her life is not only high in prudential value; it is also high in moral, aesthetic, and other value.

Not only do we use ‘is happy’ in evaluative ways; we also use it to express a *psychological*, or mental, fact about the person who has it, where ‘Sally is happy’ expresses something like the proposition that Sally is currently experiencing a particular psychological state: happiness. Thus construed, ‘is happy’ is *descriptive*, not necessarily evaluative, since there is nothing in the meaning of the term that mentions ‘value’ (or its synonyms): it is logically

possible for Sally to be happy but her life not be high in prudential value.⁷ In order to avoid unnecessary confusion, hereafter, following the literature, I shall use terms like ‘is happy’ and ‘happiness’ to connote this latter, *descriptive*, sense. And I shall use ‘well-being’ and ‘welfare’ to connote the first *evaluative* sense such that ‘well-being’ and ‘welfare’ track just *prudential* value.⁸

I want to briefly return to feeling happy. There appears to be no contradiction implied by the sentence ‘I feel happy but I am not happy’. What may be going on here is the employment of two different senses of ‘happy’. The first conjunct ‘I feel happy’ may employ the *descriptive* sense of happiness, where ‘I feel happy’ means I currently possess the psychological state: happiness. The second conjunct ‘I am not happy’ may employ the *evaluative* sense of happiness, where ‘I am not happy’ means I am not well off. Thus understood, the sentence ‘I feel happy but am not happy’ means I am happy but am not well off. While I think this makes some sense of the distinction between feeling and being happy, I am not convinced that it explains the distinction completely.

1.1.3 Being occurrently, dispositionally, and predominately happy

Distinguishing actual from apparent psychological happiness is a good start; but I also need to distinguish what it is to be occurrently happy from what it is to be dispositionally happy,

⁷ Notice that this is a logical or semantic point about the concept of happiness, not a substantial claim about whether happiness is good for the happy. In section 4.2 I discuss the metaphysical issue of whether happiness is intrinsically prudentially valuable.

⁸ Haybron also distinguishes these three different senses of ‘happiness’ (2008: 208-213). Although Feldman (2010: 8-10) and Goldstein (1973) do *not* discuss the use of ‘happiness’ to connote overall value, they discuss the distinction between ‘happiness’ used to connote a psychological state and ‘happiness’ used to connote an evaluative state.

and this involves distinguishing them both from what it is to be predominately happy.⁹ A person is *occurrently* happy because of ϕ at time t if and only if she is happy because of ϕ right at t . Occurrent happiness is experienced in *moments* or *episodes*. If you are currently happy right now, you are occurrently happy. Suppose that you are watching the Broncos play the Seahawks in the Super Bowl. If while watching the Seahawks win you are made happy by the refreshing taste of the ice-cold beverage you are drinking, you experience occurrent happiness. And if drinking a second ice-cold beverage would make you happy at some future moment t_1 , you are dispositionally happy in regard to the second beverage at t in the sense that that beverage disposes you to be happy at t_1 . A person is *dispositionally* happy because of ψ only if, other things equal, she would be made occurrently happy by ψ if she were to think about ψ or experience it.¹⁰

In addition to being occurrently or dispositionally happy, you can also be *predominately* happy over a period of time. If I were to say that you were predominately happy, I would be saying you are approximately happy, that although you may have experienced some unhappiness, you were (occurently) happy more often than not. It follows that you are predominately happy over a duration d if and only if you experience more occurrent happiness over d than occurrent unhappiness.

⁹ Feldman also discusses the distinctions between being dispositionally and predominantly happy (2010: 56-57) and occurrent and dispositional pleasure (2010: 111-112). Because he thinks that pleasure is essentially attitudinal, Feldman describes the distinction between occurrent and dispositional pleasures using propositions as objects. You have occurrent pleasure over p if the fact that p is true is pleasurable to you. You have dispositional pleasure over p when if you were to think about p , its being true would be pleasurable to you. When I use ‘occurently happy’ or ‘dispositionally happy’ I do not mean to suggest that pleasure is essentially attitudinal or even that pleasure is the essence of happiness. I think that there is a distinction being occurrently and dispositionally happy, whatever the correct theory of pleasure or happiness turns out to be.

¹⁰ Davis also mentions the distinction between being occurrently and dispositionally happy. But he defines the latter in terms of being predominantly happy: “In the dispositional sense, ‘A is happy’ means that A is predominantly happy in the occurrent sense” (Davis 1981a: 305, 1981b: 111). I agree with Feldman, who criticizes Davis for failing to use the term ‘dispositional’ in its ordinary everyday sense (Feldman 2010: 56-57). I use ‘predominantly happy’ to express Davis’s being dispositionally happy.

Occurrent happiness, then, is epistemically prior to both being dispositionally and predominantly happy. In order to understand what it is to be dispositionally or predominantly happy, we have to understand what it is to be occurrently happy; while one may understand what it is to be occurrently happy without also understanding what it is to be dispositionally or predominantly happy. I am therefore interested in explicating actual *occurrent* happiness.

1.1.4 Relational and non-relational happiness

There are at least two ways that you may be happy: you may be 1) relationally happy, or you may be 2) non-relationally happy.¹¹ You may be happy *about* something or *with* something; or you may be happy *that* something is the case; you may even be happy *to*, as in happy to be here. These are all examples of *relational* happiness, where ‘happy’ occurs with a complement. Used in this manner, ‘happy’ is a dyadic, not a monadic, term. In these cases, you are not just happy; rather, you are happy in regards to something. In some cases, your happiness has an intentional object.¹² For example, you may happy *about* or *with* a friend of your finding a job.

¹¹ Davis (1981a: 111) and Sumner (1996: 143-147) make this distinction, too. Davis distinguishes between relational happiness and non-relational happiness the same way I do. Sumner distinguishes between 1) being happy with or about something, 2) feeling happy, and 3) being happy/having a happy life. Being happy with or about something is “roughly equivalent to being satisfied or content with it” (143). Feeling happy does not require an intentional object; instead, it roughly amounts to the feeling or mood of contentment or joy (144). Being happy or having a happy life roughly consists of being positively disposed to your life (145-146).

¹² One, it seems, can be relationally happy in two distinct ways: 1) intrinsically, and 2) extrinsically. You are intrinsically happy that *p* if *p* by itself brings you happiness; whereas, you are extrinsically happy that *p* if it brings you happiness *only insofar as* it relates to something else that brings you happiness. That is, one is made happy by *p* intrinsically if and only if there is no *q* such that one is made happy by *p* in virtue of being made happy by *q*; *p* is the final object of your happiness. Whereas you are made happy by *p* extrinsically only if there is some other object *q* such that *p* makes you happy in virtue of *q* making you happy; *p* makes you happy not because it alone brings you happiness; rather, one is made happy by *p* because it brings about *q*, which by itself makes you happy (See, for example, Feldman (2004), where he distinguishes intrinsic from extrinsic *pleasure*).

Intrinsic happiness, then, is epistemically and ontologically prior to extrinsic happiness. In order to understand extrinsic happiness, we need to first understand intrinsic happiness. And this is because you are extrinsically happy that *q* only if you think *q* brings about *p* and you are intrinsically happy that *p*. I previously said that I am interested in both non-relational and relational happiness. When it comes to relational happiness, I am primarily interested in intrinsic, not extrinsic, happiness.

In other cases, your happiness has a propositional object. For example, you may be happy *that* Barrack Obama won re-election. There is also non-relational occurrent happiness, where ‘happy’ occurs *without* a compliment. Used in this manner, ‘happy’ is a monadic, not a dyadic, term. In such cases you are not happy in regards to, related to, something; you are just happy. Non-relational happiness, then, has no intentional or propositional object.

Suppose that you have a craving for your favorite cigar. Unfortunately, you have not been able to acquire one in years; your local cigar store always seems to be sold out of them. But the day finally arrives: the store now has a few in stock. As you start smoking the cigar, you are reminded by how incredible the taste is. Smoking this cigar makes you happy. Moreover, as you enjoy smoking the cigar, you also find yourself happy that you are smoking the cigar.

In this example you experience both non-relational and relational happiness. Smoking the cigar causes you to be happy. And you are also happy *that* you are smoking such a fine cigar. The former is a token of non-relational happiness while the latter is a token of relational happiness. In the former case, while smoking the cigar caused you to be happy, there is no intentional or propositional object that you are happy over; you are just happy. In the latter case, there is a propositional object that you are happy in regards to: the state of affairs of you smoking your favorite cigar. I see *no* reason for supposing that relational happiness is different in kind from non-relational happiness; rather, they seem to be the same *type* of phenomenon: it just so happens that one has a propositional or intentional object, and the other does not.

Some argue that relational happiness is essentially attitudinal, that being relationally happy essentially amounts to having the relevant pro-attitude toward the object of your happiness. Now, there seem to be many different pro-attitudes, like: admiration, approval, enjoyment, hope, intending, liking, preference, satisfaction, and want, amongst others. So if

relational happiness consists of having a pro-attitude, what particular pro-attitude does it consist of? Well, it seems obvious that you can be happy over φ without admiring φ , hoping that φ , or preferring φ . And while it may not seem obvious that you can be happy over φ without approving of, liking, or wanting φ , all three pro-attitudes seem to be some form of satisfaction. Not only does it not seem obvious that you can be happy over φ without enjoying φ or be satisfied that φ , neither enjoyment nor satisfaction obviously entails the other. Satisfaction and enjoyment, then, are plausible answers for what pro-attitude relational happiness consists of, if it consists of an attitude at all.

Those who claim relational happiness consists of satisfaction maintain that “[b]eing happy with something is roughly equivalent to being satisfied or content with it” (Sumner 1996: 143-144). Relational happiness, understood in such a way, is a cognitive matter with no necessary affective component: you may be happy that something is true without experiencing any occurrent feeling. Those who think that relational happiness consists of *enjoyment*, not satisfaction, maintain that being happy with something is roughly equivalent to enjoying the fact that something is true. Enjoyment differs from satisfaction in two relevant ways. First, you may enjoy p without being satisfied p is true. Second, enjoying p may entail that you occurrently feel something in regards to p .¹³ Point being: relational happiness understood as satisfaction would make relational happiness just a cognitive matter, whereas relational happiness understood as enjoyment leaves open the possibility that relational happiness is affective.

For my money, relational happiness is best understood as enjoyment. I am interested in both non-relational happiness and relational happiness. I will argue (in chapter three) that pleasure is the essence of both non-relational happiness and relational happiness. It just so

¹³ A lot is riding on how we understand ‘enjoyment’. Some argue that S’s enjoying the fact p is true does *not* imply that enjoyment consists of an occurrent feeling (See Feldman, 2004 and 2010). My point is: conceptually speaking, enjoyment can be understood in such a way as to entail an occurrent feeling whereas mere satisfaction cannot.

happens that the pleasure associated with relational happiness has intentional or propositional content whereas the pleasure associated with non-relational happiness does not. I think that we commonly use ‘enjoyment’ to refer to pleasure that has intentional or propositional content. That is how I will be using it in the rest of the dissertation.

Now that I have said a few things about happiness, I need to discuss what I mean when I say that I am engaging in a *philosophical* analysis of happiness.

1.1.5 Philosophical analysis of happiness

This *philosophical* project does not concern itself with the likely causes of happiness. Being in love, raising a family, owning vast material wealth, whatever you think likely causes happiness, so be it; just recognize that this does not provide us an answer to our question: what *is* happiness? I am also *not* interested in practical advice on how to get happiness. Investing in blue chip stocks, going overseas as an Oxfam volunteer, marrying young, whatever you think will best enable you to achieve happiness, so be it; just recognize that this also does not get at the essence of happiness. For that, I need to construct, as Feldman states, a “suitably general statement of necessary and sufficient conditions” of what it is for a person to be happy (Feldman 2004: 13).

So I attempt to construct such a statement in chapter three of this dissertation. If I am right, hedonism about happiness is the correct theory of happiness. That is, if I am correct, very roughly, happiness is nothing essentially other than the experience of pleasure. This is not, however, to say that every time we say something like ‘Sally is happy’ we mean something like Sally is experiencing more pleasure than pain. People use ‘happy’ to express all sorts of things, some not very precise, after all. My thesis, HH, is a metaphysical project about the phenomenon of happiness. This is distinct from the linguistic project about the meaning of ‘happy’. My aim

is to discover what it is that all and only happy people have in common, not necessarily what ‘happiness’ means.

Now, there may be nothing common among all phenomena we call ‘happy’. Our use of ‘happiness’ may be arbitrary or at least vary to such an extent that makes it impossible to perform a philosophical analysis of. While this may be right, the only way to know if ‘happiness’ is not capable of being analyzed is by first trying to analyze it. To presume that happiness is incapable of being analyzed in this manner without even trying to do so is clearly rash and therefore unjustified. Let us try to analyze it, then. If it turns out we cannot do so, then we cannot do so; but it may turn out that we can.

It is also possible that the best we can do with the term ‘happiness’ is label it a “family resemblance” concept. Terms like ‘game’ or ‘sport’ are sometimes said to bear family resemblance in that although no statement of necessary and sufficient conditions of the essence of a game or a sport can be made intelligible, there are sufficient similarities between all phenomena described by ‘game’ or ‘sport’ such that we recognize what ‘game’ or ‘sport’ mean; it is just that those similarities are impossible of being summarized in a formula. Perhaps our efforts will reveal that while we may gain a better understanding of happiness by discussing some similarities many (but not all) happy people have in common, we cannot declare with confidence that happiness has an essence. But again, the only way to justifiably believe that being happy is a family resemblance property is by first trying to explicate happiness. Clearly, it would be irrational to presume that being happy is a family resemblance property without first trying to describe what it is to be happy.

Not only is analyzing happiness a philosophical project, I contend that it is a philosophical project *worth* pursuing. While there is nothing in the concept of happiness that makes it an evaluative concept, I shall argue later in section 4.2 that happiness is good for the happy: whenever it is present, it seems to make us better off. Indeed, the theory of welfare that I defend in chapter six implies that happiness is a constituent of welfare. If this is correct, clearly, happiness is worth investigating.

I have yet to address a pertinent question: if I am interested in the descriptive, psychological, sense of ‘happy’, why is this a philosophical project? Is it not the domain of *psychology* to analyze the *psychological* sense of happiness? I think not. When we philosophers analyze a phenomenon, we attempt to give an account of that phenomenon’s essence. A *philosophical* analysis of happiness consists of describing what it is that all and only happy people have in common. This is a task suited for philosophers since we have the training to do it; it is what we do. Psychologists are more interested in empirical issues like measuring how happy people are, helping people attain happiness, and examining what happiness looks like on a brain scan. They are less interested in describing what all and only happy have in common than they are in describing how happy a person appears to be. Knowing what the psychological literature says about happiness is useful for understanding what it essentially is; I am not disputing that. But surely we may use philosophical methods and see how far they get us. We can try to determine the plausibility of a given theory of happiness by weighing it against our intuitions and by determining whether there are any internal problems within the theory. If what follows is correct, we can use these methods to arrive at the correct theory of happiness.

This concludes my discussion of preliminary matters. For the remainder of this chapter I introduce five examples of happiness.

1.2 Examples of happiness

I mean for the following cases to be uncontroversial examples of happiness. The people I describe, our non-inferential prereflective inclinations tell us, are happy. The correct theory of happiness should accord with our commonsense intuitions on these cases, or in the event it does not accord, provide us with an adequate explanation for why we are mistaken.

1.2.1 Brainwashed Sally

This is the case from the introduction. Over the past few months Sally feels great. Never has she been so confident of the truth of her convictions, socially adept with others, and just fulfilled. And not only does she experience much pleasure, her feeling so much pleasure pleases her, too. Although she feels so great, Sally is a member of destructive cult. She would not be so high on life if she were not being brainwashed. Fortunately for Sally, her parents see what is really going on and conduct an intervention on her behalf. The intervention goes well, so well that Sally leaves the cult. In the first month or two after she leaves Sally often feels miserable and depressed. She is deeply unsatisfied with her lot. Yet she is now beginning to exercise her critical capacities in a manner she did not while brainwashed. And she takes pleasure in relying on her own beliefs and capacities rather than those of another. Several years later Sally is no longer struggling with her departure from the group. She often feels good, just not as good as when she was brainwashed. Sally is pleased that she left the cult, so pleased that she would rather continue her life outside the group than return to it. Further, she is now a more rational person than she was even prior to joining the cult.

1.2.2 Mary the mother¹⁴

Mary is giving birth to her first child, which she always wanted. While giving birth she experiences a wide variety of emotions, from elation to dread. She also feels intense pain: the drugs the doctors give her can only do so much. Yet, she enjoys giving birth to her child; indeed, she has so much enjoyment that she would describe this *moment* as the happiest of her life thus far.

1.2.3 Glenda the graduate student¹⁵

Glenda is a graduate student who above all both wants to write a great dissertation and obtain a tenure-track position at a research university; but she is *not* confident that she will satisfy either desire. Perhaps Glenda doubts herself because she does not possess the requisite skills to excel in her academic discipline: she is just a moderately gifted graduate student. Glenda toils day after day on her dissertation with no apparent prospect of success. This toiling weighs her down, leaving her feeling despondent and anxious most of the time. Her depression prevents Glenda from sleeping more than a few hours a night. Not only does she often feel bad, not too many people enjoy her company; indeed, most would describe her in one word: ‘irritable’. Fortunately, Glenda is now seeing a psychiatrist.

The psychiatrist first tries to counsel Glenda by applying cognitive behavior therapy. Her psychiatrist encourages her (despite Glenda’s only moderate talents) to believe both that she will write a great dissertation and procure a tenure-track position at a prestigious university. However, no matter how hard the psychiatrist tries, Glenda refuses to give up her beliefs; they are just too entrenched. Accepting Glenda’s obstinacy, the psychiatrist now attempts to change

¹⁴ I take this example from Feldman (2010: 33-34).

¹⁵ I also take this example from Feldman, who calls Glenda ‘Susan’ (2010: 63-65).

her desires. This, too, ends in failure: her desires are like her beliefs—too entrenched for her to give up. So the psychiatrist changes his tactics. Instead of trying to get Glenda to change her beliefs or desires, the psychiatrist now prescribes some mood altering medication, medication that is intended to make the pessimistic more cheerful. Thankfully, the medication works as advertised, improving Glenda's mood considerably: although she still has the same beliefs and unsatisfied desires, Glenda is no longer so pessimistic. She feels cheerful, not depressed. Glenda now sleeps well at night, enjoys working on her dissertation, and is much less irritable.

1.2.4 Seneca the Stoic

Being a Stoic, Seneca has unusual self-control, learning long ago not to permit his emotions from interfering with his reason. He exercises this self-control by not allowing himself to be emotional. Imagine that he has just given an excellent oration on why the Roman Empire needs to adopt fiscal and judicial reforms whereby slaves would be treated more humanly. Seneca is pleased with himself and his speech: it seems likely that Rome will act in the manner he prescribed. Yet if you were to look at Seneca in the moments following his oration you would not be able to tell how much pleasure he experiences. He does not seem particularly cheerful: not even a smile cracks his lips. Indeed, his demeanor is the same it would have been if Seneca had not just given the speech. He is *no* more cheerful or optimistic than he was before the speech. Nonetheless, he has a lot of enjoyment right now, not much displeasure.

1.2.5 Andrew the swine

Andrew's father is a farmer. But being that his dad has enough hired hands, Andrew is left with much free time. So he spends his life in utter debauchery, using illegal substances and engaging in illicit activities like fornicating with some of his father's hired hands and even committing bestiality with the farm animals. Why does Andrew engage in such debauchery?—

the answer is simple: because he receives a lot of pleasure from doing so. Now, as you can imagine, Andrew is not the most reflective person; he just enjoys the feelings his activities give him. During his more lucid moments, Andrew is displeased by the fact that such activities feel so pleasant to him. Now, this displeasure does *not* come close to the strength or intensity of the pleasure he gets from the activities themselves. Overall, Andrew is a very cheerful person. He just happens to enjoy what we would likely say are disgusting activities.

The above five cases, I contend, are all uncontroversial examples of happiness. Prereflectively I think that we are inclined to judge the people described as happy. In the next chapter I consider whether three popular theories of happiness accord with them and whether the theories have internal problems.

Chapter 2: What Happiness is Not

In this chapter I discuss three different theories of happiness. After explicating each, I discuss to what extent the theory accords with intuition in the five examples of happiness discussed above. If I am correct, each theory fails to accord with intuition in at least one case; this is because, I argue, the things these theories claim constitute happiness do not in fact constitute happiness. I start with desire-satisfaction about happiness.

2.1 Desire-satisfaction about happiness (DSH)

2.1.1 The essence of DSH

Roughly, DSH is the position that happiness is the satisfaction of your desires: you are happy to the extent that your desires are satisfied, unhappy to the extent that they are frustrated. It is important to note how strong a position DSH is: it implies that the satisfaction of *any* of your desires makes you happier, period. Now, in order to understand DSH we have to understand what proponents of DSH mean by both ‘desire’ and by ‘satisfaction’.

A desire is a mental state; more precisely, it is an attitude, not necessarily a feeling or an emotion. A desire is not just any attitude, however: it is a *pro-attitude*. To desire something is to be positively disposed to that thing. There is a long-standing tradition that understands desiring ϕ to be nothing essentially besides being motivated to do ϕ . Like belief, then, while you may be aware of a desire you have, you need not be. You seem to desire many different things without being vividly aware that you do, namely: breathing, hunger, thirst, and so on. Hence, proponents of a desire view acknowledge that the extent to which a particular desire-satisfaction contributes to happiness is determined, at least in part, by how *intense* and how *long* the desire is. Basically, the more intensely and the longer you desire ϕ the more satisfying your desire will increase your

happiness. Desires are similar to beliefs in at least one further way: there is propositional content to a desire; whenever you desire something you want the content of your desire to obtain.

Suppose that you desire to be with your family at Christmas. You would then desire the proposition that you are with your family at Christmas be true. Now that I have provided a brief account of what desires are, let us move on to satisfaction.

There are two different ways to understand ‘satisfaction’. First, we may understand ‘satisfaction’ as *objective* satisfaction, where a desire for ϕ is satisfied if and only if ϕ . Thus construed, DSH maintains that you are happy to the extent that you get what you want, unhappy to the extent that you do not.¹⁶ Now, it could be the case that you get what you want but do not know about it; then again, it could also be the case that you do *not* get what you want but you nevertheless believe that you do. The point is: for objective DSH, it is not whether you believe that you get what you want that constitutes happiness; it is actually getting what you want that constitutes happiness.

Second, we may understand ‘satisfaction’ as *subjective* satisfaction, where a desire is satisfied if and only if the person who has the desire *believes* that the desire is satisfied. Thus construed, DSH maintains that you are happy to the extent you believe you get what you want, unhappy to the extent you do not.¹⁷ According to subjective DSH, you may get what you want or you may not; either way, it does not constitute happiness; the only thing that constitutes your

¹⁶ V. J. McGill is sometimes referred to as a proponent of *objective* DSH (See Feldman 2010: 53). I, however, am not comfortable referring to McGill as a proponent of DSH. True, he does espouse a desire-satisfaction theory, but he seems to offer it as a theory of *welfare*, not happiness. Like Aristotle, he considers happiness to be the highest good we can achieve, that which is “(1) desired for its own sake, and (2) not desired for the sake of anything else, and (3) will be the only good of which (1) and (2) are both true” (McGill 1967b: 13). Although he does not distinguish between psychological and evaluative happiness, McGill seems to be using ‘happiness’ to express something like the value-laden concept of welfare, not the value-neutral psychological state I refer to as ‘happiness’.

¹⁷ For perhaps the most sophisticated versions of subjective DSH, refer to Davis (1981b) and Chris Heathwood (2005 and 2007); although Heathwood does not explicitly maintain that satisfied desires constitute happiness, he constructs a sophisticated formulation of desire-satisfaction about welfare that employs this subjective sense of ‘satisfaction’.

happiness is your *believing* that you get what you want. Belief, like happiness, may be *occurrent* or *dispositional*; proponents of DSH maintain that both are relevant for your happiness. You need not be occurrently thinking about a desire that you believe has been satisfied in order for it to increase your happiness.¹⁸ You believe that $2+2 = 4$, but you may not have thought about it in months. Presumably, you have beliefs that you are not completely aware of and beliefs while you are sleeping; these would be dispositional beliefs.

According to DSH you are happy to the extent that your desires are satisfied; the extent to which a particular desire is satisfied contributes to your happiness is determined by the desire's intensity and duration. But remember: happiness is a psychological or mental state that you experience in moments *at a time*. Your happiness at a particular time seems to be determined by happiness-constitutive experiences or states that you have at that time. DSH may seem to fail to accommodate this restraint. Suppose that you intensely desire to run a marathon in the year 2016. But between now and then, say, in 2015, you stop desiring to run that marathon. When 2016 rolls around, however, circumstances dictate that you run the marathon, anyway.¹⁹ Does running the marathon make you any happier? If so, when does it make you happier?—back in 2014 when you desired to run that marathon, or in 2016 when you actually ran it?

Neither of these answers is correct. Your psychological state back in 2014 is not at all affected by what you do in 2016; so how can running a marathon in 2016 affect how happy you are back in 2014? And in 2016 when you actually run the marathon, you do not even desire to run it, however; so according to DSH, you are made no happier by running it. The obvious answer to the question is no: satisfying that desire makes you no happier. DSH can

¹⁸ See Davis (1981b: 113).

¹⁹ Dorsey (2013: 153-154) uses a similar example in his defense of a desire-satisfaction theory of welfare.

accommodate this if it includes a *concurrency* restraint such that ϕ is a satisfied desire for S at a time t *only if* S desires ϕ at t .²⁰ Thus formulated, only satisfied desires that you desire at the time they occur constitute your happiness. DSH, modified in such a way, maintains that you are happy at t to the extent that your desires at t are satisfied at t . Going forward, I shall understand DSH as this formulation.

2.1.2 DSH and our five cases

Let us now run both types of DSH through the gauntlet of the above five cases, starting with Sally. Although we are inclined to believe that Sally is happy both while she is being brainwashed and several years after she leaves the brainwashing, she seems happier brainwashed: she is more cheerful, feels better about herself, has more pleasure, and more satisfied desires brainwashed than not. We are also inclined to believe that Sally is unhappy in the first month after she leaves the cult because she often feels downright miserable. What is DSH's verdict?

Well, according to *subjective* DSH, Sally is happy to the extent she believes that her desires are satisfied, unhappy to the extent she believes they are not. Over the months Sally is brainwashed she believes that she has satisfied many of her desires, and she believes she has failed to satisfy few of her desires; indeed, over this duration Sally has more net subjective desire-satisfaction than during any other similar duration in her life. Thus, subjective DSH deems Sally happy, the happiest she has ever been (over a similar duration). Years after she leaves the cult Sally believes that she is satisfying more desires than not, but Sally does not believe she is satisfying as many desires as she was while she was being brainwashed; and even though she believes that she fails to satisfy few desires years after the brainwashing, she believed

²⁰ Dorsey (2013) and Heathwood (2005) defend a desire-satisfaction theory of welfare with a concurrency constraint.

that she failed to satisfy even fewer desires while brainwashed. According to subjective DSH, then, Sally is happier brainwashed than years afterward. Subjective DSH also deems Sally unhappy in the first couple months she leaves the brainwashing because she does not believe that she satisfied many of her more intense desires. Hence, subjective DSH accords with our intuitions in the case of Sally. The same goes for *objective* DSH: objective DSH deems Sally happy while brainwashed because more of her desires go fulfilled than unfulfilled; and objective DSH implies that Sally is happy years after she leaves the cult, just not as happy as she was in it because not as many of her desires are fulfilled as they were previously (and more go unfulfilled). Objective DSH also deems Sally unhappy in the first month after the intervention, since many of her desires go unfulfilled. Hence, no matter how we interpret ‘satisfaction’ DSH accords with our pretheoretical inclinations about Sally.

It seems reasonable to suppose that both subjective and objective DSH also accord with the intuition that Mary is happy while giving birth to her first child. Sure, she experiences much pain; but both versions of DSH deny that pain intrinsically affects how happy Mary is. What matters, according to DSH, is that she intensely wants to give birth to her child, she believes that she is satisfying that desire, and she is in fact satisfying that desire. That said, Mary does *not* want to experience the pain of childbirth. Sure, she may be said to desire it *extrinsically* in the sense that what she desires most of all is to give birth and she therefore desires anything necessary—including the pain of childbirth—to fulfill that desire. But proponents of DSH usually formulate DSH such that the fulfillment of only *intrinsic* desires makes one happier.²¹ Since Mary does *not* desire the pain of childbirth *for its own sake*, her pain, then, does not contribute to her happiness. So, for the proponent of DSH, whether Mary is happy comes down to this: whether she prefers giving birth more than preferring not being in pain. Since she prefers

²¹ See both Davis (1981a and 1981b) and Heathwood (2005 and 2007).

the former over the latter and she believes that she is giving birth, subjective DSH deems her happy. Likewise, since she prefers giving birth to not experiencing the pain, objective DSH deems her happy, too.

The case of Glenda poses a problem for DSH. Remember, Glenda is the one who above all else desires to write an excellent dissertation and ascertain an excellent professorship. While many of her unconscious desires go satisfied—like breathing, eating, drinking, and so on, her *most intense* desires do not: she neither believes that she writes an excellent dissertation nor does she actually write an excellent dissertation; neither does she believe she obtains a tenure-track position nor does she actually obtain one. That said, her mood changes due to the medication her psychiatrist prescribes such that although her desires remain unsatisfied, she no longer feels despondent or anxious; rather, she feels great, enjoying long bouts of optimism and cheerfulness. This is why Glenda seems happy, or at least happier than prior to taking the medicine. But both versions of DSH deny that Glenda is any happier: after she takes the medication, although her mood improves, she is still not getting what she wants; nor does she believe that she is getting what she wants. Thus, both subjective and objective DSH deem Glenda just as happy while depressed as not depressed. This seems wrong, however: Glenda seems happier medicated and cheerful than unmedicated and depressed.

The proponent of DSH has a response to this. She could argue that Glenda believes that more of her desires go satisfied after she takes the medication and that more of Glenda's desires are actually satisfied after taking the medication. She feels cheerful after taking the medication, not despondent or anxious. Obviously, she desires to be cheerful, at least more so than she desires to be despondent or anxious. And because more of Glenda's desires go satisfied (both subjectively and objectively), DSH deems her happier after the medication. This is *not* a good

response. DSH does not deem the satisfaction of any desire as equally happiness-constitutive as any other; rather, the extent to which a particular satisfied desire is happiness-constitutive is determined by how intense and how long the desire is. In the case of Glenda, while she desires to be cheerful and does not desire to be depressed, the intensity and length for which she desires being cheerful pales in comparison to her most intense and long-lasting desires of writing an excellent dissertation and obtaining a tenure-track position. According to DSH, then, roughly speaking, Glenda is no happier after taking the medication, which seems wrong.

Although they fail to accord with intuition in the case of Glenda, both versions of DSH accord with intuition in the cases of Seneca and Andrew: Seneca and Andrew both believe that they are doing what they want and both get what they want. In the case of Andrew, he wants to do illegal drugs; he wants to fornicate with his father’s hired hands and commit bestiality; doing so brings him much pleasure, after all. While it is also true that upon reflection Andrew desires not to have such base desires, he does not reflect all that often: so he has more base desires than desires not to have such desires, and his desires for base pleasures are much stronger than his desires not to engage in such activities. Andrew, then, both believes that he gets much of he wants and he actually gets much of what he wants. Hence, both forms of DSH deem Andrew happy, which accords with our prereflective inclinations.

In the table below I sum up the results of to the extent to which DSH accords with intuition in the five cases.

	Intuition	Subjective DS	Objective DS
1. Brainwashed Sally	H	H	H
2. Mary the mother	H	H	H
3. Glenda the talented student	H	U	U
4. Seneca the Stoic	H	H	H

5. Andrew the swine	H	H	H
---------------------	---	---	---

2.1.3 Additional considerations

Judging from the above cases DSH may seem like a plausible theory of happiness; it accords with intuition in four of the five cases, after all. I contend, however, that appearances are deceiving: desire-satisfaction seems wrong about happiness. *Objective* DSH, in particular, seems wildly implausible. I shall criticize it first, then move on to criticizing *subjective* DSH.

Suppose that you desire ϕ . It seems strange that ϕ 's obtaining can make you happy without you believing that ϕ has obtained or without you gaining pleasure from ϕ 's obtaining. Indeed, if *objective* DSH is correct, ϕ makes you happier even if it does not affect your mental states. Suppose that you are travelling on a subway and notice someone, call her 'Sarah', on crutches with a cast around her leg.²² Being that you are a morally decent person, you desire for Sarah to get well, for her leg to heal. Suppose that weeks later, unbeknownst to you, Sarah's leg heals. When it is healed, a desire of yours has been fulfilled. Thus, according to objective DSH, you are happier in virtue of Sarah's leg being healed, even though you do not know it has been healed; but how can this be? Sure, Sarah may feel happier; her leg is no longer broken, after all. But how does the mere fact that her leg has healed affect your psychology such that you are happier?

The above case strikes us as strange, I contend, for a reason: I think that *objective* DSH is in tension with itself. Remember, I am using the term 'happiness' to denote a descriptive state of your psychology, of your mental states. So happiness is a *mental state* or a series of mental states. If something affects your happiness, it must affect a mental state of yours. But if

²² I borrow this example from Parfit (1984: 494).

objective DSH is correct, your happiness may increase even if your mental states do not change. This is not possible, however. If the theory is not incoherent, there is at least a considerable tension that needs to be resolved.²³

The above criticism seems reason enough to render objective DSH an implausible theory of happiness. Now I shall criticize *subjective* DSH. My criticism of subjective DSH relies on cases like Glenda: she seems happier after taking the medication even though she believes that her intense desires are unsatisfied. Having your most intense desires satisfied just does not seem necessary for you to be happy. Not only that, satisfying intense desires seems *insufficient* for happiness. Think of Daniel, an ascetic. He deeply desires to commune with God; so he joins a monastery. Such an environment gives him plenty of time to reflect on God's mysterious ways. The thing is: much of the time he feels deep melancholy. Now, he knows that he could just leave the monastery to feel better; but Daniel does not want that: he prefers the melancholy that comes from his service to God to the pleasure of a more "worldly" existence. Even though he believes that he gets what he yearns for, he does *not* seem happy.

Not only does desire-satisfaction seem neither necessary nor sufficient for happiness, some satisfied desires seem to leave us unhappy, not happy.²⁴ Take Oscar. For months now he has yearned for the release of *Dungeon Siege II*. Indeed, not a day goes by where he is not intensely looking forward to playing this video game. Once it is released, however, and he begins playing it he is sorely disappointed: it is just not fun to play, or at least not nearly as fun to play as he expected it would be. Now, Oscar does not feel disappointed as soon as begins playing but five minutes into it; call this moment 't⁵'. He gives the game a few days before giving up and shelving it for the foreseeable future. According to DSH, since his intense desire

²³ Using a similar argument Sumner deems *objective* theories of happiness "unintelligible" (1996: 140).

²⁴ Such desires are what Sidgwick calls 'dead sea apples' (1981: 110).

for playing the game has been satisfied (both subjectively and objectively), he should be thereby happier. Yet he feels disappointed and thereby less happy. Oscar desires to play the game at t^5 and he believes that he is playing the game at t^5 ; yet he feels unhappy at t^5 .

There is a fix to this problem, but it requires a substantial change to DSH. Rather than maintain that anything you desire affects your happiness so long as it occurs at the time you desire it, DSH could maintain that *only sensations that you experience at the same time that you desire them (for their own sake)* make you happier. While you desire to play that video game, playing the game is not an experience for you. The *disappointment* is, though. But because you do not desire the disappointment when you are experiencing it, it does *not* make you any happier.

While this modified DSH is no longer susceptible to the worry I described above, this modification has considerable cost: we no longer have a desire-satisfaction theory of happiness, but something similar to my own theory of happiness. What kind of experiences are you going to desire for their own sake? The obvious answer is *pleasant* ones. Many philosophers contend that there is nothing essential to a pleasant experience besides you desiring that experience for its own sake when it is occurring.²⁵ If pleasure is understood in this way, not only is my own theory of happiness coextensive with this modified DSH, this modified DSH just *collapses* into my own theory.

My last criticism of subjective DSW concerns the desire to be unhappy. It may sound strange, but it seems possible for us to desire our own *unhappiness*. It is not hard to think of individuals who want to be unhappy; Daniel the ascetic, described above, is one example. Yet it is difficult to think of individuals who want to have their desires frustrated. As Feldman says, “this consideration by itself may be sufficient to show that happiness cannot be identified with

²⁵ In section 3.1.1.1 below I discuss this theory of pleasure along with another; I also provide a list of the philosophers who espouse a particular theory.

desire satisfaction” (2010: 68), or at least that unhappiness cannot be identified with desire frustration. I want to add some more substance to this important criticism; to do so, I shall use the example of Daniel. Suppose that he only has one intense desire at t : to be unhappy at t . Call this desire for unhappiness at ‘U’. He either believes that he is happy at t , or he believes that he is unhappy at t . On the one hand, if he believes that he is *unhappy* at t , according to subjective DSH, he is *happy* at t : he desires U and he believes U; that is, his intense desire has been satisfied; which makes him, according to DSH, happy. On the other, if he believes that he is happy at t , according to subjective DSH, he is unhappy at t : he desires U and he believes U; that is, his intense desire has been frustrated, which makes him unhappy. This is not quite paradoxical, but it is odd that no matter what Daniel believes about his happiness he is mistaken.²⁶

Now that I have criticized DSH, let us move on to our next theory of happiness.

2.2 Life-satisfaction about happiness (LSH)

2.2.1 The essence of LSH

LSH is the position that happiness is the satisfaction of our lives as a whole: you are happy to the extent that you are satisfied with your life as a whole, unhappy to the extent that you are dissatisfied with it.²⁷ This does not merely involve satisfaction with one’s life; it involves satisfaction with one’s *whole* life. By ‘whole life’ most advocates of LSH mean a person’s life up to the moment in question.²⁸ The most plausible LSH theories consider satisfaction to consist

²⁶ Feldman makes the same point (2010: 68).

²⁷ Proponents of LSH include Benditt (1974), Montague (1966), Sumner (1996), and Tatarkiewicz (1976).

²⁸ A minority take ‘whole life’ to mean past, present, and future life; see Tatarkiewicz (1976: 140).

of two components, one cognitive, the other affective. A person is *cognitively* satisfied with her life as a whole to the extent her life on balance at least roughly “measures up favorably against [her] standards or evaluations” (Sumner 1996: 145-146). A person is *affectively* satisfied with her life as a whole to the extent that she experiences “a sense of well-being: finding [her] life enriching or rewarding, or feeling satisfied or fulfilled by it” (Ibid.).²⁹ A person, then, is not happy according to LSH unless she is generally speaking both cognitively *and* affectively favorably disposed to her whole life: cognitive satisfaction and affective satisfaction are each necessary for happiness; together they are sufficient, or so says LSH.

Remember how there are two ways of understanding ‘satisfaction’ in desire-satisfaction theories of happiness?—well, the same is true for LSH. On the one hand, we may understand ‘satisfaction’ to mean *actual* satisfaction, meaning you are happy if and only if you are actually overall satisfied with your whole life. On the other hand, we may understand ‘satisfaction’ to mean *hypothetical* satisfaction, meaning you are happy if and only if *if* you were to think about your whole life, you would be overall satisfied with it. Hypothetical satisfaction is *less* stringent than actual satisfaction. According to the former, you need *not* be occurrently reflecting about your life and find it to be on the whole worthwhile. No, you only need to find your life on the whole worthwhile in the counterfactual where you are actively thinking about your life. Actual satisfaction is more stringent: in order for you to be happy, you must be occurrently reflecting about your life and find it on the whole satisfactory. Now, neither proponents of actual LSH nor proponents of hypothetical LSH maintain that you must be completely satisfied with your whole

²⁹ Proponents of LSH are in rough agreement on what they mean by ‘satisfaction’. See, Richard Brandt, who says that a person is *cognitively* satisfied with her life when she “does not want [it] to be substantially different, and that [it] measure[s] up, at least roughly, to the life ideal [she] had hoped to attain” (1967: 413-414). He also says that a person is *affectively* satisfied with her life when she “feel[s] joy or enthusiasm or enjoy[s] what [s]he was doing or experiencing” (Ibid.).

life for you to be happy. Rather, they merely require that you be *on the whole* satisfied with your life: there may be parts of your life that you are not satisfied with; but so long as you are satisfied with your whole life overall, LSH deems you happy.

We can measure how happy you are over a period of time or how happy you are at a single moment. Your *momentary* happiness, if LSH is right, is nothing other than how affectively satisfied you are at that particular moment and how cognitively satisfied you are with your life up to and including that moment. Your happiness over a *duration* is simply the sum of how happy you are over each moment of that duration.

Now that I have explicated LSH, it is time to consider how well it accords with intuition in our five cases.

2.2.2 LSH and our five cases

Let us first recall Sally, who feels great, gets much pleasure, and satisfies many of her desires even though she is being brainwashed. *Actual* LSH deems that Sally is happy so long as she is satisfied with her life as a whole, while *hypothetical* LSH deems Sally happy so long as if she were to reflect about her life as a whole, she would be satisfied with it. Sally is clearly *affectively* satisfied while she is brainwashed: she feels fulfilled, the most fulfilled she ever has.³⁰ Sally is also *cognitively*—actually and hypothetically—satisfied with her life. While she is displeased by the fact that she wasted the vast majority of her life thus far, Sally regards her life prior to the joining cult as necessary to getting her to where she is today. Sure, she would prefer it if she always had her new beliefs—that is, prefer to have always been a member of the group;

³⁰ Some critics of LSH fail to discuss affective satisfaction in their criticisms of LSH. I am not sure why this is when proponents of LSH are clear that satisfaction includes an affective dimension. Examples of such forgetful critics are Feldman (2010: 83-90) and Haybron (2008: 82-96).

but she recognizes that her prior life was necessary for her to be a member of the group and thus is overall satisfied with it.

Both versions of LSH also accord with the intuition that Sally is unhappy in the first month after she leaves the cult. She feels far less fulfilled outside the cult than inside; in fact, she feels downright lousy. Cognitively, Sally is not at all satisfied with her life, either: she is coming to grips with the fact that she wasted a year of her life following the whims of a megalomaniac. Years after she leaves the brainwashing Sally feels much better: she generally feels cheerful and optimistic and usually experiences more pleasure than displeasure. Sally also fares better cognitively: now that she has learned about the nature of brainwashing, she understands what it was about her that enabled someone else to take advantage of her. With the aid of therapy, Sally tends to reflect a lot about her life. She has thus come to accept that she was brainwashed and is now even attempting to learn from it. That said, she was more affectively and cognitively satisfied while brainwashed. Thus, according to either version of LSH, Sally is happy years after she leaves the brainwashing, not happy a month after the brainwashing. And both actual and hypothetical LSH deem Sally happiest while brainwashed. LSH, then, accords with intuition in the case of Sally.

This brings us to our next case: Mary. Mary is not currently reflecting about her life. She does not form something like the judgment *I am satisfied with my life thus far*: Mary is too busy giving birth to engage in such reflection. Actual LSH, hence, deems her unhappy at the moment she is giving birth. And while Mary finds her current experience of giving birth rewarding and fulfilling, *if* Mary was to reflect about her life as a whole, she would *not* find it roughly satisfactory. You see, Mary is a single parent whose life thus far falls far short of the expectations and values she has possessed since childhood. Hypothetical LSH, then, deems her

unhappy, as well. Hence, both actual and hypothetical LSH conflict with intuition in the case of Mary. Mary clearly seems happy when she gives birth; she even claims that she is the happiest she has ever been.

But the proponent of LSH is not done, yet. While LSH fails to deem Mary happy at the moment she is giving birth, it could be modified to deem Mary happy over the duration in which she is giving birth. The problem with LSH with this case is that while Mary is (both cognitively and affectively) satisfied with her life over the *duration of giving birth to her child*, Mary is not roughly satisfied with the *duration of her life to date*. According to LSH, you are happy at moment m to the extent that: 1) you are affectively satisfied at m , and 2) you are on balance cognitively satisfied with your life up to and including m . Thus, LSH fails to deem Mary happy.

Let us modify LSH to the position that you are happy over m to the extent that 1) you are affectively satisfied at m , and 2) you are cognitively satisfied with your life as it is at m .

Modified in such a way, you are cognitively satisfied at m to the extent that your life at m meets your expectations and standards at m . To determine how happy Mary is over the duration of her giving birth, we just add up all the moments of life-satisfaction she has over that duration. If we do this under modified LSH, Mary is happy: she feels fulfilled during most of the moments of that duration and she is cognitively satisfied with her life over that duration. This modification, then, works in the case of Mary; it does not, however, in the case of Glenda.

Intuition tells us Glenda is happier after taking the medication, since it leaves her feeling uplifted and optimistic instead of despondent and pessimistic. So she meets the affective dimension of satisfaction for LSH. Before she takes the medication Glenda reflects about her life fairly often; she is a graduate student, after all. What she finds out is *not* satisfactory to her: she views much of her life as a failure because she thinks that she has fruitlessly spent a good chunk of her

life pursuing a career she does not have the abilities for. Hence, actual LSH deems her unhappy before she takes the medication. So far, so good. But after she takes the medication Glenda does *not* reflect about her life terribly often: she forms no judgment either that she is or is not satisfied with it. Actual LSH, then, deems Glenda unhappy after she takes the medication, in conflict with intuition. Hypothetical LSH conflicts with intuition, here, too: if after taking the medication Glenda were to reflect about her life, she would be just as dissatisfied with it as she was previously. Hypothetical LSH, then, deems Glenda unhappy after she takes the medication, too, in conflict with intuition. Modified LSH also conflicts with intuition. At any particular moment over the duration in question Glenda is not only not cognitively satisfied with her life to date, she is not satisfied with it at that particular moment. It does not meet her expectations; it just so happens that she does not reflect about it all that often.

The proponent of LSH may have a response to this. The proponent could state that while Glenda is neither hypothetically nor actually cognitively satisfied, she does feel as though she is leading a fulfilled and rewarding life. This at least makes her *feel* happy. Although we would be wrong to say that she is happy, we would be right to assert that she feels happy. We started discussing the possible distinction between being happy and feeling happy back in section 1.1. LSH provides us with a way to understand the distinction. The proponent of LSH may maintain that while being happy has both cognitive and affective components, feeling happy only has the latter.³¹

³¹ See Sumner: “The cognitive component of happiness is therefore beyond the range of many subjects-of-a-life, such as small children and non-human animals. However, there is more involved in being happy than being disposed to think that your life is going (or has gone) well. The affective side of happiness consists in what we commonly call a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it. Because it is less cognitively demanding than a judgment about how one’s life is going as a whole, it is what we have in mind when say that a child or an animal is happy, or is leading a happy life” (1996: 146).

While this response allows LSH to better accord with intuition in the case of Glenda, it still fails to completely accord with intuition. Remember, Glenda just seems happy after she takes the medication. Sure, common sense suggests that she feels happy after taking her medication; but it does not suggest, like LSH does, that she is *not* happy. That is, we are not inclined to say that she *merely* feels happy.

In the case of Seneca, while it is true that after he finishes his oration he has no judgment that his whole life is worthwhile, if he were to reflect about his whole life, he would find it satisfactory; he is Seneca, after all. He also does find his life rewarding and fulfilling; he just does not show it. Thus while actual LSH deems him unhappy, hypothetical LSH accords with intuition by deeming him happy.

Both versions of LSH clash with intuition in the case of Andrew. Although he enjoys using illegal substances, fornicating with farm hands, and engaging in bestiality, he fails to reflect about his whole life. He has the necessary cognitive capacities, of course; he is just not a reflective person. Hence actual LSH deems him unhappy. And if he were to reflect about his life, Andrew would find it deeply unsatisfactory: he would feel ashamed of enjoying the pleasures that he does; perhaps this is why he does not reflect all that often. Hence, although we are prereflectively inclined to judge Andrew happy, neither version of LSH deems him happy. Modified LSH also deems Andrew not happy. Not only would he not be cognitively satisfied with his life to date, he would not be cognitively satisfied with his life at any particular moment over the duration in question.

Again, the proponent of LSH could maintain that while Andrew is not in fact happy, he nonetheless feels happy. This makes the view more palatable, but prevents it from fully

according with intuition, since, contra LSH, Andrew not only merely seem to feel happy; he also seems happy.

In the table below I sum up the results of to the extent to which LSH accords with intuition in the five cases.

	Intuition	Actual LSH	Hypothetical LSH
1. Brainwashed Sally	H	H	H
2. Mary the mother	H	U	U
3. Glenda the grad student	H	U	U
4. Seneca the Stoic	H	U	H
5. Andrew the swine	H	U	U

2.2.3 Additional considerations

Unlike DSH, LSH maintains that there are two components of happiness: an affective one and a cognitive one. I doubt that either component constitutes happiness. I shall examine the affective component first. Of course I think that feelings of well-being, fulfillment, enrichment, and the like contribute to happiness; I just wonder whether feelings of this sort by themselves contribute to happiness. Sumner admits that his theory of happiness implies that small children and non-human animals are not capable of being (un)happy, since they lack the cognitive capacities for life-satisfaction. They may, however, *feel* happy: they may experience feelings of well-being, fulfillment, enrichment and such. So Sumner is committed to the position that one may experience feelings of well-being and such without your life actually being well off, fulfilled, or enriched.

I wonder why Sumner focuses on these feelings in particular. I suppose Sumner could say that it is just a brute fact that these feelings constitute happiness; it is not as if these feelings have anything in common; they just constitute feeling happy, and that is all there is to it. At

some point, explanations need to bottom out, and this is it. I am skeptical of this retort: I think Sumner would be appealing to a brute fact too early. I agree that at some point explanations must bottom out; I just doubt that Sumner is there, yet.

Try to divorce the feelings of well-being, fulfillment, enrichment and the like from any pleasure. Imagine them no longer as pleasurable experiences. It is difficult to do, right? I think this is for a reason: there is something about these feelings that seems necessarily pleasant. If you feel well off, fulfilled, enriched, there is something pleasant about them. So while I admit that these feelings contribute to happiness, I think they contribute to happiness *only in virtue of* being pleasurable. Thus, the *pleasure*, not the feelings themselves, is the more fundamental constituent of happiness.

Not only do I doubt that the feelings LSH maintains constitute feeling happy by themselves contribute to feeling happy, I also question the cognitive element of LSH. It is easy to see that *hypothetical* LSH accords with more intuitions than *actual* LSH. Actual LSH just seems to set too high a standard for what constitutes happiness: having the judgment that your life is satisfying at moment *m* does not seem necessary for you to be happy at *m*. It seems that one can be happy without occurrently reflecting in such a manner. Although hypothetical LSH fares better, it only accords with intuition in two cases. As actual LSH seems to set too high a standard for what constitutes happiness, so does hypothetical: having the hypothetical judgment that your life is satisfying at *m* does not seem necessary for you to be happy at *m*.

LSH does not seem to be an adequate account for what it is for a person to be occurrently happy. While the happy may often be satisfied with their lives to date, it seems possible for them not to be. LSH seems to conflate *being happy with your life* with *being happy*.³² It seems

³² Carson (382-383) makes a similar point.

possible for one to be happy without also being happy with one's life. The above examples of Mary, Glenda, and Andrew are cases in point. If this is correct, then, the relationship between happiness and life-satisfaction seems merely contingent; LSH, thus, seems false.

2.3 The emotional state theory of happiness (EH)

2.3.1 The essence of EH

EH is the position that the essence of happiness is psychic affirmation, *emotionally* responding to your life as a favorable one:³³ you are happy to the extent that you are favorably emotionally disposed your life, unhappy to the extent you are negatively emotionally disposed to your life.³⁴ EH should strike us as similar to LSH, which requires one to be both cognitively and affectively favorably disposed to one's life. EH just breaks the conjunct in half, tossing the notion of cognitive satisfaction aside.³⁵ According EH, happiness concerns the *psyche*; so LSH goes wrong in focusing both on cognition and one's overall emotional condition: happiness merely concerns the latter, not the former. And your overall emotional condition has two

³³ Daniel Haybron is the author of EH. He says that we commonly distinguish between plain happiness and true happiness, where true happiness is reserved for states where we feel we could not be leading a more fulfilling life. Such a state, Haybron says, is not merely psychic affirmation, but psychic flourishing, when our lives are "more or less" perfectly matched to our emotional nature (2008a: 148-149). Using Haybron's term I am interested in plain happiness, *not* necessarily true happiness.

³⁴ Haybron's EH is more nuanced than my presentation of it. Although he does maintain that happiness is constituted by psychic affirmation, emotionally responding to your life as a favorable one, he argues that there are three modes of emotional response: attunement, engagement, and endorsement (2008a: 147). I am not convinced these three modes are relevant for our purposes, however. My explication of EH differs from Haybron's in at least one other respect: Haybron does not commit himself to the claim that one is happy *to the extent* that one's emotional condition is positive, unhappy *to the extent* one's emotional condition is negative. Rather, he maintains that if one is happy, one's emotional condition is *broadly* positive, experiencing only minor negative emotional responses at most (Ibid.).

³⁵ Haybron makes a similar point in comparison between LSH and EH (2008a: 111). But he ignores the affective dimension to LSH and insists that whereas LSH concerns being cognitively favorably disposed to one's life, EH concerns being affectively favorably disposed to one's life. This is an incorrect dichotomy to make, since LSH includes an affective dimension, too.

components: 1) your central affective states, and 2) your mood propensities (Haybron 2008a: 147). We need to discuss both; I shall start with central affective states.

EH distinguishes between two different degrees of affective states: *central* and *periphery*. The former “get to us” whereas the latter “bounce right off us” and are thereby quickly forgotten (Haybron 2008a: 129). Because happiness gets to us, it is composed of central affective states, *not* peripheral ones. Central affective states “constitute changes in us”, whereas peripheral affective states are “merely things that happen to us” (Haybron 2008a: 130). Further, central affective states *dispose* us “to experience certain affective states rather than others” (Ibid.), whereas peripheral states do not. A depressed mood, for example, tends to make us pessimistic, causing us to experience little pleasure and look on the darker side of things. Anxiety, for another example, tends to make us “multiply and exaggerate potential threats” (Ibid.).³⁶ The mild annoyance someone feels after dropping her pencil, however, and the trivial pleasure one receives from scratching a nagging itch are examples of *peripheral* affective states. Sure, you may have an emotional reaction to the dropping of a pencil; but your annoyance is only brief and fleeting: it bounces right off you. Likewise for the pleasure you receive from scratching an itch: although the pleasure is an emotional reaction, it is merely a peripheral affective state since it does not dispose you to experience other affective states; although it impacts you, the impact is fleeting.

Now that we discussed the central affective states aspect of EH, let us move on to its other aspect: *mood propensities*. Happiness, Haybron says, does not merely consist of positive

³⁶ Central affective states, Haybron tells us, also tend to possess four other characteristics. First, they are productive, causing many and varied consequences, like the creation of another affective state or a physiological change (2008a: 130). Second, central affective states are persistent in that they have a long duration; they do not vanish quickly as peripheral states do (Ibid.). Third, they tend to pervasively “permeate the whole of consciousness” (Ibid.). Fourth, central affective states tend to be profound in that they deeply affect us in a way mere physical pleasures do not (2008a: 131).

central affective states; it also consists of something “deeper and more continuous”: your “propensity to respond [...] favorably to things” (2008a: 135). Think of someone who experiences “a preponderance of positive affect result[ing] purely from a fortunate confluence of positive events” and compare her to someone who experiences positive affect not by chance but “from an underlying endogenous condition” (Ibid.). The latter, not the former, Haybron contends, is a lasting condition, just like happiness. While moods are longer lasting than emotions, they are not as long lasting as temperaments or personalities. Your personality is how you are “characteristically disposed to react”, while your temperament “var[ies] considerably over time” (Haybron 2008a: 137).

And propensities are dispositions: possessing a positive mood propensity is to be disposed to respond to things in a certain way. Haybron provides us with an example of a friend who “has had a hair-trigger propensity for anxiety in recent months” (2008a: 135). The mere fact that your friend has this propensity, Haybron contends, makes her less happy than she would have been without the propensity.³⁷ Although not as essential to happiness as affective states, moods are relevant. Moods have this *tie-breaking* contribution to happiness: if two people experience the same amounts of positive and negative affective states but one has a more positive mood propensity (after subtracting any negative mood propensity) than the other, that person is thereby happier than the other.

2.3.2 EH and our five cases

EH certainly deems Sally happy, both during the brainwashing and several years later. According to EH, Sally is happy to the extent that she has positive central affective states and to

³⁷ Haybron is careful to note that some mood propensities seem irrelevant to happiness (2008a: 137). Suppose that you are disposed to be mildly annoyed by the dropping of a pencil. Although you possess this propensity, it does not seem relevant to your (un)happiness. The propensities that are relevant for happiness, Haybron maintains, are those that are *emotionally-based*: they directly affect your emotional state (Ibid.).

a lesser degree, to the extent she has positive mood propensities. While being brainwashed, Sally experiences many more positive affective states than negative: she usually feels optimistic, cheerful, confident, and powerful. A Month or two after leaving the cult Sally is struggling with the loss she feels. She often feels depressed and despondent, not at all cheerful; indeed, this is a miserable time for her. She thus experiences more negative affective states than positive, for which EH deems her unhappy, in accordance with intuition. Although she experiences fewer positive affective states years after she leaves the brainwashing, she still has more positive affective states than negatives ones. Thus, while EH deems that Sally is happy during and years after the brainwashing, it also deems her happier brainwashed, in accordance with intuition.

So long as the joy Mary experiences from giving birth to her first child is experienced in a central affective state, EH also deems Mary happy. True, she is experiencing much pain, so much pain that it is unlikely that it does not get to her. Her pain, then, appears to be a central affective state, not a peripheral one. But so long as her joy outweighs her pain, which it does, EH deems Mary happy, in accordance with intuition.

I want to change the scenario. Suppose that Mary's pain outweighs her joy; consequently, while Mary still thinks of herself as happy when she gives birth, she no longer considers it her happiest moment. In this modified case, EH no longer deems Mary happy: Mary's emotional condition is now mostly negative. And this may constitute a problem for EH: intuition seems to deem Mary happy in this latter case, too; because EH does not, EH seems mistaken.

It is not difficult for the proponent of EH to respond to this. EH could be modified such that rather than implying that Mary is happy to the extent that she has positive affective states, it implies that Mary is happy if she is in a broadly positive affective state. Because Mary is in

neither a broadly positive or negative state, this modified EH deems her neither happy nor unhappy. While it still conflicts with intuition since it fails to deem Mary happy, it does *not*, like unmodified EH, deem Mary unhappy. For my money, I do not think that this modification is necessary.

First, note that while intuition is clear in the original case that Mary seems happy, intuition is not as clear in the modified case. Indeed, if we supposed that Mary's pain far outweighs her joy, we would *not* deem her happy even if she thought that she was. We also have some reasons to be skeptical of the infallibility of self-reports of happiness in general.

According to Haybron, we appear to have a bias in favor of viewing ourselves as happy (2008a: 212). Think of "Polyannas" who always look on the bright side of things or "Kvetches" who just focus on their negative experiences. There is some empirical evidence that seems to confirm this bias. A study by Lykken and Tellegen for Western countries reveals that 86 percent of subjects tested "rated themselves more "happy and contented" than about two-thirds of the population" (qtd. in 2008a: 212). And a survey completed in 1995 by the University of Michigan found that 94 percent of Americans reported being happy (qtd. in 2008a: 216). It does not seem plausible that so many of us are happy.

Not only do we have reasons for doubting the infallibility of self-reports of happiness in general, we also have some reasons for doubting that someone in Mary's situation will be infallible about reporting her happiness. Basically, Mary may have very good reasons for thinking of herself as happy. Here are four. First, perhaps it is in her best interest to believe that she is happy (even though she is not). Or, secondly, it could be Mary's motherly instinct kicking in: her believing that she is happy may benefit her child. Third, sometimes we find it difficult to come to grips with our actual affective state at a given time and instead rely on what we should

feel like in that situation. Haybron has us think of a groom on his wedding day (2008a: 209). Rather than “come to grips with the complexity of his emotions”, he thinks of himself as “filled with pure joy, mainly because that’s how he is supposed to feel, and how he imagines any newlywed must feel” (Ibid.). Perhaps Mary thinks of herself as happy because that is how a mother is supposed to feel when she has just given birth. Fourth, it even seems possible that we could be mistaken about the valence of our emotional state (Haybron 2008a: 209-210). Refer back to the groom; suppose that he is experiencing more negative affect than positive. The thing is: he sincerely believes that he is experiencing more positive affect. He is a bridegroom on his wedding day, after all; and grooms are supposed to be happy.

Let us now proceed by considering how well EH accords with the cases of Glenda, Seneca, and Andrew.

EH certainly deems Glenda happy. After taking the medication she feels great, uplifted rather than despondent. The medication she takes is designed to improve her mood, and that is exactly what it does. Although she remains skeptical of her academic talents and although her desires remain unsatisfied, she has many more positive affective states than negative ones. So EH accords with intuition here, as well.

Seneca has just delivered an impressive oration in front of the Senate. Although he is pleased by this fact, he does not show any outward signs of it; he is a Stoic, after all. From appearances, Seneca seems neither happy nor unhappy, just calm. No smile or frown adorns his face. He learned to suppress his emotions long ago. Seneca, then, is not currently experiencing any emotional states, let alone positive ones. Nor is he inclined to be in moods in the future, let alone positive ones. He is neither disposed to be cheerful or anxious, jubilant or depressed but just affectively neutral. Because he is not experiencing emotions, EH fails to deem him any

happier after delivering his oration than before. But he is experiencing pleasure over delivering a good speech, and he is experiencing no pain. Intuition tells us that he is happy, at least happier after giving the oration than before, contra EH.

Of course Haybron could respond by asserting that it is not possible for Seneca to experience pleasure without also experiencing that pleasure in an emotional state. This seems wrong, however. While there is considerable debate over the nature of pleasure, almost everyone agrees that pleasure is essentially either some sort of sensation, an attitude, or a combination of both sensation and attitude. I of course admit that some pleasures are emotional pleasures; but I also contend that there are many pleasures that involve *no* emotion. Think of the pleasurable warmth of a hot shower on a cold winter day. While the pleasurable warmth you feel is certainly a sensation, it does not appear to be an emotion. Or think of the pleasure you may have over Seahawks winning the Super Bowl. You may express this pleasure in an emotional outburst when the game finishes; but it seems that you may enjoy the win without expressing it in an emotion, too.

In the table below I sum up the results of to the extent to which EH accords with intuition in the five cases.

	Intuition	EH
1. Brainwashed Sally	H	H
2. Mary the mother	H	H
3. Glenda the talented	H	H
4. Seneca the Stoic	H	U
5. Andrew the swine	H	H

2.3.3 Additional considerations

While I think that the case of Seneca is sufficient to render EH implausible, I want to discuss in detail why neither component of EH—central affective states or mood propensities—seems to constitute happiness. Let us start with mood propensities, then discuss affective states.

Haybron asserts that mood propensities contribute to happiness merely in a tie-breaking sense. In the event that two individuals share the same level of positive central affective states EH looks at their moods to determine who is happier. But mood propensities are dispositional, *not* occurrent. While it is true that if two individuals S and U share the same level of occurrent happiness at t_1 but S is disposed to be happier than U at t_2 , then S may be said to be *dispositionally* happier than U at t_1 , she is not thereby any occurrently happier than U at t_1 . Just because an individual is dispositionally happier than another at a certain time does not mean that she is also occurrently happier at that time. Mood propensities, because of their dispositional nature, then, do not seem to be an essential part of *occurrent* happiness.³⁸

Haybron responds to criticisms like mine by offering three arguments against them. He first suggests that Sumner depicts LSH as *primarily* dispositional: “to be satisfied with one’s life, e.g., just is, at least in part, to be disposed to think about one’s life in certain ways, to have certain feelings when contemplating one’s life, and so forth” (Haybron 2011: 27). To support this reading Haybron refers to Sumner, 1996: 146. This reading misunderstands Sumner, however. On the previous page (145) Sumner clearly distinguishes what he calls ‘having a

³⁸ Scott Hill (2007) also criticizes Haybron for including mood propensities as a component of happiness; but Hill gives them too much importance: Hill characterizes EH as stating that mood propensities are a necessary condition for happiness, whereas Haybron gives them something like tie-breaking importance (see Haybron (2011: 21), where Haybron clarifies the role of mood propensities in his theory). Haybron (2011) attempts to defend EH against Hill by arguing that EH accords with intuition in the two cases Hill contends that EH does not accord with. I do not wish to step into that debate since Hill’s criticism relies on a mischaracterization of EH. My criticism, however, relies on no such mischaracterization. Morris’s (2011: 269) criticism of EH, though not identical, is more like mine: he argues that “Dispositions only have extrinsic as opposed to intrinsic value so far as happiness is concerned. Any value they confer derives from the experiences they help to bring about”. Hence, Morris concludes, dispositions “are not an essential component of happiness” (Ibid.).

happy disposition/personality’ from what he calls ‘being happy/having a happy life’, the latter of which he is “principally concerned”. His description of the *cognitive* aspect of LSH follows.

The cognitive aspect of happiness consists in a positive evaluation of the conditions of your life, a judgment that, at least on balance, it measures up favorably against your standards or expectations. This evaluation may be global, covering all of the important sectors of your life, or it might focus on one in particular (your work, say, or your family). In either case it represents an affirmation or endorsement of some (or all of) the conditions or circumstances of your life, a judgment that, on balance and taking everything into account, your life is going well for you. (145)

Now, Sumner occasionally slips into using language that may suggest that LSH includes a dispositional element: “there is more involved in being happy than being disposed to think your life is going (or has gone) well” (this is the bit Haybron (2011: 27) cites); but charitably interpreted, it is clear that Sumner does not mean for LSH to consist of a dispositional element, let alone a *primary* dispositional element, as Haybron suggests. This interpretive argument that Haybron gives, then, falls flat: Sumner does *not* seem to intend for LSH to be interpreted as primarily dispositional.

Let us suppose that Haybron is right, however. That is, let us suppose that Sumner understands LSH as being primarily dispositional. What is the upshot?—what does this do for Haybron? My criticism that dispositional elements—like Haybron’s mood propensities—do not belong in a formulation about the essence of *occurrent* happiness would now just apply to Sumner in addition to Haybron. So even if Sumner did maintain that LSH includes an essentially dispositional element, it is not clear how this addresses the worry.

Haybron relies not only an argument based on interpretation for his claim, he also argues that LSH is best understood (independent of how Sumner describes it) as dispositional: how else, Haybron states, can LSH avoid requiring one to have an “occurrent thought or feeling about her life at every moment” without referring to disposition (2011: 27-28)? The answer is simple: as described above, we understand LSH to mean *hypothetical*, not actual, satisfaction. Now, we

could understand the type of satisfaction in LSH as dispositional; I will give Haybron that. But *we need not to*: LSH seems just as plausible, if not more plausible understood hypothetically than dispositionally. After all, understood dispositionally LSH faces the criticism I have levied against Haybron.

Haybron's last argument for including a dispositional component in a theory of happiness is that, "arguably", including such a component better tracks "popular views of the unconscious" (2011: 28). Here, Haybron implies that if we were to understand happiness as merely a collection of occurrent, *conscious*, mental states, then thinkers in the psychoanalytic tradition would reject it. I am no expert in psychoanalysis; so I will not debate whether psychoanalysts have such a position. Instead, I shall suppose Haybron is right to think that it is possible for unconscious states to be a part of happiness, and thus that we do not want to endorse a theory of happiness that rules out that possibility. But I fail to see why Haybron equates occurrent states with conscious states: supposing that happiness merely consists of occurrent mental states does not imply that happiness consists of occurrent, *conscious*, mental states. I could occurrently experience a myriad of affective states and not be conscious of them. One of Haybron's own examples illustrates this. Suppose that you have the unfortunate circumstance of possessing an old refrigerator, one that whines apparently due to a bad bearing. Over time you fail to notice the racket that the machine makes. But occasionally, when the compressor stops, you notice a sudden glorious silence. You also notice that you have a painful headache, apparently from the obnoxious racket that you did not realize was occurring until now. You were in pain and just did not realize it.

While Haybron does not say this, there is a more convincing response that he could give. He could accept that there is nothing in the concept of occurrent happiness that implies a

dispositional element, but nonetheless contend that there is also nothing in the concept that proscribes a dispositional element. Happiness could be like the color red. I own a red car. Its redness is merely dispositional; it merely has the disposition for redness: under the right circumstances, to the right sort of perceivers, it appears red to them.³⁹

Let us refer back to Haybron's example. He describes a friend who "has had a hair-trigger propensity for anxiety in recent months" (2008a: 135). Refer to the friend as 'Antha'. Antha possesses a propensity for anxiety in that if the right circumstances obtained, she would be anxious. Suppose that we catch Antha on a good day, when the right circumstances have not yet obtained to trigger the anxiety. According to Haybron, we would conclude that she is less happy on this day even though she has no anxiety than on a day where she not only is not anxious but also where she is "relatively immune to anxiety" (2008a: 135).

I disagree with Haybron's intuition here. Antha's propensity for anxiety and the redness of my car are disanalogous: while the right circumstances have obtained to make my car red (when someone views it), the right circumstances have not yet obtained to make Antha anxious. If they did, they would make her anxious. In that case, her *anxiety*, not her propensity for the anxiety, seems to make her less happy. The mere fact that Antha has this propensity, Haybron contends, makes her less happy than she would have been without the propensity. This does not seem right to me. If Antha has this propensity for anxiety at t^1 but does not experience anxiety at t^1 , I fail to see how Antha's happiness at t^1 is impacted by the propensity. Sure, Antha's happiness at a later moment, say, t^3 , when the circumstances have obtained making her anxious, may be impacted by her propensity at t^1 ; but this is not what the proponent of EH needs.

³⁹ According to Haybron, mood propensities are not object-specific: you are not disposed to be in a certain mood "in response to particular objects or events"; rather, you are disposed to be in a certain mood over a wide range of circumstances (2008a: 137). So the right circumstances here have to be wide ranging.

Haybron provides more examples that we should consider. Suppose that you are at a funeral and see grieving family members of the deceased laughing with old friends. We would not, according to Haybron, “think their unhappiness has completely, if temporarily, lifted”; rather, we would think that “[a] deeper unhappiness remains” (2008a: 135-136). While I agree with Haybron that we would think of the grieving family members as unhappy despite their laughter, I doubt that we would think this because of some propensity for grieving that they are not currently exercising. Rather, a deeper unhappiness seems to remain in the family members because they are still grieving: although they are laughing with close friends and thus may seem to be happy, they are likely experiencing some dissatisfaction, pain, or negative affect at the loss of their family member; they just do not express it. Back in section 1.1.1 I discussed the distinction between merely seeming to be happy and being happy: one can display signs of happiness without being happy. Suppose that I just lost my job. My friends take me out for a few drinks to get my mind off it. While I may appear jolly in their company, a deeper dissatisfaction I have lies underneath the surface. While I may seem happy, I am really unhappy. Then again, maybe the family members experience a few moments of happiness while laughing with their close friends; over the duration of the funeral, however, the family members are *predominately* unhappy; and so a deeper unhappiness remains. My point is: in order to conclude that the grieving family members are unhappy despite their appearances, we need not adopt the idea of mood propensities.

Haybron offers another example that we should consider: the case of Tom and Jerry.

They both have a happy temperament and take a relaxing vacation to the beach.

During their several days together, their moods are generally quite similar and fairly positive, save that Tom is a bit more cheerful, being pleased finally to get away from home some months after a difficult divorce. By and large he is not the least melancholic—most of the time he feels wholly unburdened, laughter comes easily, and he takes great pleasure in catching up with his old friend and conversing with other vacationers. Yet Tom’s emotional state remains unusually

fragile: on two occasions he bursts into tears and weeps uncontrollably. These episodes don't last long, so that Tom's moods are still, on the whole, a little more positive than Jerry's. (2008a: 136)

If Haybron is right, "Tom's elevated propensity for sadness diminishes his happiness in itself, and [...] he is *less* happy than Jerry" (Ibid.). I again differ with Haybron. While I agree that Tom's sadness diminishes his happiness, I just do not think that his propensity for sadness divorced from his sadness also diminishes it. I would, then, conclude that Tom is not less happy than Jerry, contra Haybron. If, on the whole, Tom's moods are indeed more positive than Jerry's, then Tom seems, on the whole, happier. Although he is more likely than Jerry to burst into tears and experience intense sadness in a wide range of circumstances, these circumstances do not obtain except in two short bursts, which do not prevent Tom from feeling "wholly unburdened" and from feeling much pleasure most of the time. If Tom is indeed "a bit more cheerful" than Jerry over this duration, then, he seems a bit happier, too, despite his fragility.

Not only am I skeptical of EH's reliance on mood propensities, I also question whether happiness essentially consists of positive central affective states. I shall focus on centrality, then positivity. If I am correct, happiness is neither a central nor a positive affective state.

While Haybron may be right in that happiness may have more to do with one's affective states than one's cognitive states, does happiness seem to be just a matter of possessing certain emotions or moods? Let us consider what these certain emotions and moods consist of. Haybron maintains that the only emotions and moods that constitute happiness are central, not peripheral, states. *Central* affective states get to us and constitute changes in us; they also dispose to act in characteristic ways. *Peripheral* states, however, bounce right off us and hence do not get to us in the same way central affective states do. While happiness sometimes, perhaps often, does get to us and constitutes changes in us, happiness need not do so: sometimes happiness is brief and fleeting; sometimes we are happy one moment, unhappy the next. When

this occurs, happiness bounces right off us, and hence is not a central state. Suppose that you are a Seahawks fan enjoying them win the Super Bowl. While you watch the game, you feel content and you experience much pleasure, no pain. On your drive home from watching the game, however, you no longer feel so satisfied nor experience much pleasure: you have already forgotten about the game entirely; it was a pleasant distraction, but now it is time to get back to the real world. Although you seem happy while watching the Seahawks, the happiness does not seem particularly deep. While not all happiness is so shallow, some of it is. Haybron, then, is wrong to think that happiness is constituted *only* by central affective states.

I also do not think that Haybron use of ‘positive’ and ‘negative’ is justified. He labels happiness-constitutive emotions ‘positive’, unhappiness-constitutive emotions, ‘negative’. A typical positive affective state are is cheerfulness; a typical negative affective state is depression. What distinguishes positive emotions from negative ones? The natural response is: whether experiencing the emotions is *pleasurable*. Positive emotions are simply emotions that are pleasant to experience; negative emotions are simply emotions that are unpleasant to experience. Haybron disagrees with this response, insisting that it is *not* merely pleasure that distinguishes positive emotions. He says that

affective states of the same type usually, but not always, have the same hedonic properties. Just as one might find a painful sensation pleasant, so is it possible to find pleasure in anger, fear, and even sadness—say, while reading a tragedy. Yet anger, fear, and sadness look to be unhappiness-constituting whether or not they are pleasant. (2008a: 146)

Haybron’s argument, then, is something like this. If all that distinguished positive emotions was their being pleasurable, then all pleasant emotions are happiness-constituting. But not all pleasant emotions are happiness-constituting: anger, fear, and sadness are all emotions that we may find pleasure in but are nonetheless unhappiness-constituting. Thus, pleasure is not

the only thing that distinguishes positive emotions from negative ones. This is a valid argument; I just think the second premise is false.

The phrase “find pleasure in” is ambiguous. On the one hand, it may mean that negative emotions like anger, fear, or sadness *themselves* are pleasurable. On the other, it may mean taking pleasure in experiencing anger, fear, or sadness. In the former it is the experience of sadness itself that one finds pleasure in, whereas in the latter it is not the experience itself but something like the attitude one has toward the experience that is pleasurable. Haybron needs the former understanding; but all that I am willing to grant is the latter. There is nothing pleasurable in anger, fear, or sadness themselves. While we may experience pleasure while also experiencing these emotions, the pleasure is not found in the emotions themselves. At most, one may take pleasure in the state of affairs that includes one feeling angry, sad, or fearful. While reading a tragedy you may feel angry, fearful, and sad as well as take pleasure in reading the tragedy or even in feeling the way you do.

The same point applies to the examples Haybron uses.

Some people have unusual tastes or values: they *like*, find pleasant, many of the things that ordinary people find disagreeable. For instance, a melancholy person—a Keats, perhaps—might seek comfort in his sorrow, and enjoy wallowing in his own grief. For this sort of individual, his unhappiness *is* his pleasure. [...] And a tortured artist need not derive any pleasure from the cheerful affects that disgust her (feelings brought on, say, by a shamefully maudlin episode of Barney the purple dinosaur). (2008a: 72)

I do not dispute that some people find pleasure in things ordinary people find unpleasant.

Hedonism permits individuals like Keats to enjoy their own grief. Perhaps Keats finds his sorrow to enable him to write such inspiring poetry. Thus understood, it is not the grief itself that Keats finds enjoyable but rather the state of affairs where his grief is necessary for poetry. If Keats’s sorrow is more intense and lasts longer than his enjoyment, HH deems him unhappy. Imagine someone who is sad whose sadness is not at all unpleasant to her. Not only is she *not*

displeased by her being said: there is nothing about the sadness itself that is unpleasant. It is strange to think to that this person's sadness makes her unhappier. Haybron disagrees: "intuitively, a said person is unhappy *whether or not* she likes how it feels, whether or not it feels good to her" (2008a: 71). This may be simple table-pounding, but I fail to see how a person's sadness affects her happiness if the sadness is not at all unpleasant for her. And I think most people agree with me. Likewise for the tortured artist: she may be disgusted by feeling cheerful. Just because she finds her cheerfulness disgusting does not mean that she gets no pleasure from the cheerful mood.

Haybron refers to characteristically pleasant or unpleasant emotions. But what happens if we consider a more neutral emotion like surprise? Does it contribute to your happiness or to your unhappiness? The hedonist's answer is simple: it depends on whether the surprise is pleasant or unpleasant. It is not obvious what response EH would give. Indeed, Haybron contends that

it seems obvious that in some perfectly ordinary sense the happiness-constitutive states are "positive," the unhappiness-constitutive states "negative"—and, moreover, that a proper account of this can be given in scientifically respectable terms without relying on value judgments. (2008a: 146)

I admit that many emotions contribute to our happiness, like cheerfulness, joy, and euphoria; many emotions also seem to contribute to our unhappiness, like anxiety, depression and irritability. This, however, is because the emotions are either pleasant or unpleasant. An emotion, like surprise, that neither typically contributes to our happiness or unhappiness is neither positive nor negative on Haybron's account. Yet it can clearly contribute to our happiness. And if it is the pleasure, not the emotion, that differentiates positive from negative affect, the pleasure is the *more fundamental* constituent of happiness. The only emotions and moods that contribute to happiness, I contend, are pleasant ones, ones that feel good to

experience. And it is the pleasure, then, that constitutes happiness; the emotion or mood is merely the vehicle or effect of the pleasure, as far as happiness is concerned.

2.4 Chapter conclusion

In this chapter I argued against three theories of happiness: DSH, LSH, and EH. The case of Glenda seems to falsify DSH. She seems happier after taking the medication when she feels cheerful and optimistic than before taking it when she feels despondent. Yet, because she does not satisfy her intense desires before or after the medication, DSH deems her just as happy. LSH conflicts with our commonsense judgments about happiness not only in the case of Glenda, but also in the cases of Mary and Andrew. Andrew is not at all happy with the overall quality of his life; yet he seems happy. LSH seems to err in conflating happiness about your life with being happy. While EH seems more plausible than either DSH or LSH, it conflicts with intuition in the case of Seneca. One does not seem to need to be in a positive central emotional state or have positive mood propensities in order to be happy. And while our affective states do seem to contribute to our happiness, they seem to do so only in virtue of being pleasurable; and if this is correct, it is the pleasure, not the affective state, that is more fundamental to happiness.

Chapter 3: What Happiness Is

3.1 Hedonism about happiness (HH)

3.1.1 The essence of HH

In this chapter I introduce and then defend my own theory of happiness: hedonism about happiness (HH). HH is the position that the essence of happiness is pleasure: you are happy to the extent that you have pleasure, unhappy to the extent that you have pain. In order to understand HH we need to discuss both pleasure and pain. I shall start with pleasure, then proceed to pain.

Let us start off with some uncontroversial apparent facts about pleasure. Suppose that you are a Seattle Seahawks fan watching the Super Bowl. While watching the game you are drinking a crisp, refreshing, and soothingly bitter India Pale Ale, your favorite kind of beer. The aroma and flavor of the hops in the IPA make drinking it a pleasurable experience for you.

We can glean several things from this example. First, note that pleasure is an *internal* phenomenon: pleasure occurs within our minds or bodies, not in external objects. Moreover, there is always a reason for why you experience a particular pleasure, there is always something causing you to have that pleasure—in the above case, it is sensations from the drinking of the beer. Now, while there always seems to be something causing a particular pleasure, a person may be mistaken about it. Suppose that you are at a nice restaurant enjoying a glass of wine. You think that you enjoy the glass of wine because of the pleasurable feelings the wine gives you; truth be told, however, you enjoy the wine because of the fact that your dinner party sees you drinking a reputable wine.⁴⁰

⁴⁰ I take this example from Feldman (2004: 60), who uses it to make the same point.

There are plenty of other examples of pleasurable experiences; the following are but a few: the pleasurable *warmth* felt from a shower taken after shoveling deep, back-breaking, wet snow, the pleasing *taste* of a refreshing ice-cold beer after working outside in the yard all day, and the pleasant *thrill* of riding Kingda Ka, the world's tallest steel roller-coaster. The warmth you feel from the shower is an example of a tactile pleasure; the taste of the refreshing beer is a taste pleasure; and the thrill of riding Kingda Ka is an emotional pleasure. There are plenty of other pleasures: aural, gustatory, olfactory, visual, and propositional attitudinal. The last item on the list—propositional attitudinal pleasure—is *not* like the others. Where the others are feelings or sensations, propositional attitudes are not. The idea of receiving pleasure from a propositional attitude may sound strange; so a few comments are in order.

Suppose you are not only enjoying the taste of a good beer but that you are also enjoying that the Seahawks are winning. Although both instances involve pleasure, you are pleased by two different things: in the former you enjoy the *taste* of your beer, whereas in the latter you enjoy the *state of affairs* that contains your favorite team winning the Super Bowl. Your positive attitude toward the Seahawks winning brings you pleasure.

Now, pleasure caused by sensations and pleasure over states of affairs are closely related such that one often follows the other. Suppose that you have just finished shoveling, wet, deep, back breaking snow from your driveway. As a reward for your hard work, you treat yourself to a delicious, steaming, cup of hot chocolate. As soon as you begin drinking it, a feeling of warmth permeates throughout your body. You enjoy this feeling: the warmth brings you much pleasure; it is an intense, pleasurable, *feeling*. But suppose that you do not just receive pleasure from the hot chocolate; you also enjoy being in the state of receiving pleasurable sensations from the coco. In this case you have pleasure both from a sensation and from a state of affairs obtaining.

The two need not, however, always intertwine: it seems possible for you to enjoy a state of affairs without having pleasure from a sensation. Suppose that you have an arthritic grandfather, Kenneth. Due to his arthritis he usually experiences, say, 100 units of pain a day. Unfortunately, Kenneth usually experiences far less pleasure in a given day, say, 50 units. Luckily for him, however, he starts taking new arthritic medication that reduces the intensity of his pain from 100 to 20. Now, your grandfather takes much pleasure in this state of affairs, namely, that he feels much less pain than normal. Although he currently experiences no pleasure from sensations, he does take pleasure in experiencing far less arthritic pain than usual. In this case, then, we have someone who enjoys a state of affairs without also receiving pleasure from any sensation.

Likewise, it seems that you may experience pleasure from a sensation without also having pleasure from a propositional attitude. Think back to that hot chocolate you had after shoveling snow from your driveway. It seems that you *can* receive pleasure from the hot chocolate without also finding enjoyment in receiving the pleasure. While the two need not intertwine, pleasure from sensations and pleasure from states of affairs seem to be the same *type* of phenomenon; they just have different objects. Now that we have discussed pleasure in some detail, I shall discuss pain.

While hedonists about happiness all use ‘pleasure’ to describe what constitutes happiness, they differ on the term they use to describe what constitutes unhappiness. Feldman uses ‘displeasure’. But as Stuart Rachels notes, ‘displeasure’ seems to connote disapproval: if ϕ causes you displeasure, you seem to bear the negative *attitude* of being displeased that ϕ (2004: 247-248). ‘Pleasure’, Rachels claims, is not like this: it seems possible for ϕ to bring you pleasure without you being pleased that ϕ . That is, ‘displeasure’ suggests an *externalist*

understanding of pleasure such that pleasure essentially consists of an attitude. While externalism about pleasure may be true, ‘pleasure’ itself is neutral.⁴¹ Externalists should argue for their position, not “cook the books” beforehand by using a term that suggests their theory is right. I am in agreement with Rachels: ‘displeasure’ is more of an antonym for being pleased by something than it is for ‘pleasure’.

After ‘displeasure’, the most commonly used antonym for ‘pleasure’ is ‘pain’. Rachels is concerned about ‘pain’, too, however. He worries that the term excludes unpleasant states like anxiety, humiliation, and terror (Rachels 2004: 248). We do *not* seem to be in pain when we experience such states, but they are nonetheless unpleasant. Rachels’s solution is to borrow ‘unpleasure’ from the psychoanalysts and use it as the antonym to ‘pleasure’. While I think that Rachels’s point is a good one, rather than use a neologism like ‘unpleasure’, I prefer ‘pain’. But by ‘pain’ I mean to include unpleasant states like anxiety and humiliation; that is, I use the term *broadly* to include all unpleasant phenomenon.

Like pleasure, pain is an internal phenomenon, something that occurs within us. And while something always causes us to experience pain, we may be mistaken about what that cause is. Just as there are a variety of pleasurable experiences, there are also a variety of painful ones. There are tactile, taste, aural, and other pains. You may also take pain over some state of affairs. It strikes me that you may feel pain from a sensation without being pained by some state of affairs. I have often found myself feeling pain from the soreness I feel after a workout yet not distressed that I was feeling the pain. Likewise, I contend that you may find a particular state of affairs unpleasant and yet not feel pain from a sensation. You may find it unpleasant that the gap between the most well off and the worst off has never been greater in the United States without also receiving pain from a sensation.

⁴¹ I will discuss externalism and internalism about pleasure in more detail below in section 3.1.1.1.

The extent to which a particular pleasant experience contributes to happiness, according to HH, is determined by the pleasure's *intensity* and *duration*: the more intense and the more lasting an episode of pleasure is, the happier it makes you. Likewise with pain: the more intense and the longer an episode of pain is, the unhappier it makes you. HH permits both happiness in a moment and happiness over a duration. You are happy, according to HH, at a given moment m to the extent that you experience more pleasure than pain (taking into account the intensity and duration of the experiences) at m . You are happy over a duration to the extent that you experience more pleasure than pain (again, taking into account the intensity and duration of the experiences) over all of the moments in that duration.

The distinction between momentary happiness and happiness over a duration helps explain how people sometimes seem to take pride in their misery or suffering. If you are unhappy over a given duration, according to HH, you experience more pain than pleasure over that duration. While you may take pleasure in being unhappy, HH does not thereby deem you unhappy, unless the pleasure you have from your propositional attitude outweighs the pain. In that event, HH deems you happy, but only because you have more pleasure than pain. This may seem strange, but I do not think that it is all that uncommon.

Similarly, you may also be displeased about being happy. Suppose that you are eating a finely cooked meal at the establishment of your choice. While eating you reflect over how well your life is going for you: you have a modest income and are in the process of raising a family. You are well on track for satisfying many of your lifetime goals. Not to mention, the delicious taste of the meal is pleasurable to you; so is the state of affairs that includes you doing reasonably well. But then you reflect about how worse off many others are. You consider what you have done to deserve your happiness. Without being able to find an answer, you feel

ashamed of your happiness: you feel unpleasant. Now, so long as the pain does not outweigh the pleasure, HH still deems you happy. But if the pain you have from your shame outweighs the pleasure you are experiencing, HH deems you unhappy. This seems to jive well with intuition.

Relational happiness, then, according to HH, is nothing besides pleasure over an object, whether it be intentional or propositional. When you are happy that the Seahawks won the Super Bowl, you are simply taking pleasure over their having won the Super Bowl. *Non-relational* happiness, according to HH, is nothing besides pleasure; there is no intentional or propositional object that one has happiness over; rather, one is just receiving pleasure. Drinking the beer during the Super Bowl brings you pleasure; it is not that you have pleasure over your drinking of the beer; no, the taste of the beer gives you pleasure, period.

Before moving on to see how well HH fares with our five cases I would to discuss how HH is different from other hedonistic theories; and to do that I need to discuss the essence of pleasure.

3.1.1.1 The essence of pleasure

Broadly speaking, there are two different theories about what it is that all and only experiences that are pleasurable have in common: 1) *internalist* theories that postulate something internal to the experience—like a sensation or a feeling—is common amongst all pleasurable experiences, and 2) *externalist* theories that postulate something external to the experience—like an attitude—is common amongst all pleasurable experiences.⁴² If we asked the everyday person

⁴² What is said in this section about the essence of pleasure, also applies to pain: as there are internalist and externalist theories of pleasure, there are also internalist and externalist theories of pain. Presumably, if internalism about pleasure is correct, so will internalism about pain be correct. As I am neutral in regards to the correct theory of pleasure, I am also neutral in regards to the correct theory of pain: whatever pain turns out to be, it is the essence of unhappiness, or so implies HH.

on the street, I think that, generally speaking, she is inclined to say that pleasure is a kind of sensation or feeling, not necessarily a kind of attitude.

Although internalism may seem more intuitive than externalism, it faces a challenge: the so-called heterogeneity problem.⁴³ Think back to the pleasurable *warmth* felt from a shower taken after shoveling deep, back-breaking, wet snow and the pleasing *taste* of a refreshing ice-cold beer after working outside in the yard all day—what do these experiences have in common? If pleasure has an essence, there is something that these experiences share. But what common feeling or sensation is there? And we have not even discussed pleasure over states of affairs. Think back to your enjoying the Seahawks winning the Super Bowl. What possible feeling does this experience have in common with the pleasurable warmth from a hot shower or the refreshing taste of an ice-cold beverage? Pleasurable experiences just seem to be too heterogenous for there to be a common feeling or sensation that is shared amongst them.

In response to the heterogeneity problem internalists still insist that pleasure is essentially a feeling or a sensation; they just maintain that the feeling or sensation is sometimes miniscule and therefore difficult to detect. Where internalists disagree is over what precisely that shared feeling or sensation is. Some maintain that pleasure is a *hedonic tone*;⁴⁴ others maintain that it is a *distinctive feeling*.⁴⁵ The difference between the two is supposed to be that according to the hedonic tone theory, pleasure is not a particular sensation or feeling (as it is according to the distinctive feeling theory) but a *quality*, like feeling good or being enjoyable.

⁴³ See, amongst others, Sumner (1996: 92-93) and Feldman (2004: 79-81) on the heterogeneity problem.

⁴⁴ The following endorse a hedonic tone theory of pleasure: Crisp (2006), Kagan (1992), and Smuts (2010).

⁴⁵ Bramble (2011) endorses a distinctive feeling theory of pleasure.

Externalist theories of pleasure have an easy way out of the heterogeneity problem: they just deny that there is some common *internal* phenomenon to all pleasurable experiences.⁴⁶ Instead, they maintain that whatever experience brings us pleasure—whether it is the warmth of a shower or the fact that our favorite team is winning the Super Bowl—brings us pleasure because we have a positive *attitude* toward that experience. The most common externalist theories of pleasure maintain that desire is the pro-attitude that constitutes pleasure. Roughly, a *desiderative* theory of pleasure maintains that mental state *m* is pleasurable for person S to the extent that S desires *m* to occur. The most sophisticated desiderative externalist theories of pleasure are more nuanced: they maintain that S finds *m* to be pleasurable at time *t* if and only if S desires, intrinsically and de re, at *t*, of *m* that it be occurring at *t*.⁴⁷ Thus construed, it is not the occurrence of any desired mental state that is pleasurable; rather, only the desire for a mental state contemporaneous with the desired mental state is pleasurable.

One theory of happiness, the continuation theory of happiness (CH), is nothing more than a desiderative theory of pleasure combined with HH. According to CH, S is happy to the extent she desires her current experiences to continue.⁴⁸ If it turns out that pleasure is essentially desire-based, then, CH will be identical with HH. This is why rather than discussing CH as an additional theory of happiness I mention it here.

⁴⁶ See Heathwood (2007) and Sidgwick (1981: 127) for endorsements of an externalist understanding of pleasure.

⁴⁷ Heathwood endorses this formulation (2007: 32).

⁴⁸ The most well-known continuationist is Daniel Kahneman, who is regarded as the “founding father” of the new field “positive psychology” and who is also considered a founding father of happiness studies. Although Kahneman appears to offer different theories of happiness in a number of his works, he espouses a continuationist approach in his (1999) “Objective Happiness”. But in this paper he seems less interested in offering an analysis of happiness than he is in offering a stipulative definition of ‘happiness’. While Kahneman favors a continuationist approach in “Objective Happiness”, I do not mean to imply that he is a proponent of CH: CH is a theory of what it is that all and only happy people have in common; it does *not* purport to offer a stipulative definition of ‘happy’.

There is much debate about which theory of pleasure is correct. Rather than get bogged down adjudicating between internalism and externalism, I remain *neutral*. Where other hedonistic theories adopt a particular understanding of pleasure, HH does not: it just maintains that happiness consists of pleasure. Whether pleasure turns out to be a sensation or feeling, an attitude, a combination of the two, or something else, does not matter: pleasure remains the essence of happiness; or so says HH.

Now that I have discussed the essence of pleasure, it is time to determine how well HH accords with our commonsense intuitions about happiness.

3.1.2 HH and our five cases

HH deems Sally happy, both while brainwashed and several years after the brainwashing: she has more pleasure than pain—both from sensations and from states of affairs—while brainwashed and several years after leaving the cult. But because she has less net pleasure over pain after she leaves the cult than while she is a member, HH deems Sally happier in the cult. During the first month after Sally leaves the group HH deems her unhappy, since she experiences more pain than pleasure. Thus, HH accords with intuition in the case of Sally.

HH also deems Mary happy: although she experiences intense pain while giving birth to her first child, so long as her pleasure over giving birth to her first child outweighs that pain, HH deems her happy. And because her pleasure over giving birth outweighs the pain of childbirth, HH, then, deems her happy.

As we did above for EH, let us change the case. Mary still thinks of herself as happy, just not the happiest she has ever been; and instead of supposing that Mary's joy over being a mother outweighs her pain while giving birth, suppose that her pain outweighs her joy. What verdict does HH now give? Like EH, HH no longer deems Mary happy: she no longer has more

pleasure than pain; so HH deems her unhappy. Yet we would still deem Mary happy; hence, HH seems false.

I have two responses to this, each of which mirrors the responses that I provided for EH above in section 2.3.2. First, I question whether intuition is so clear on this modified case. Sure, it seems clear that we intuitively agree that Mary is happy when she is giving birth to her child: not only does she think it is the happiest moment of her life, her pleasure from being a mother outweighs the pain she experiences while giving birth. Is it so clear that we would agree with Mary that she is happy if her pain outweighs her pleasure? I doubt it. But if we did, I think that the intuition is much less strong than it is in the original case. Second, I agree with Haybron's reasons that I explicated above for being skeptical that someone in Mary's situation should be considered infallible in regards to her happiness. Mary has good reasons for thinking of herself as happy. She is, after all, giving birth to her first child; and mothers are supposed to be happy. There are plenty of explanations for why Mary thinks of herself as happy even though she is experiencing more pain than pleasure. I think that I have provided enough to make us doubt the veracity of Mary's self-report.

I should also address another criticism, which requires me to change the case of Mary yet again. Suppose that the intensity of her pleasure over giving birth to her first child is slightly less than the pain she experiences at the moment she gives birth; Mary still deems this the happiest moment of her life, however. HH now entails that Mary is unhappy: she has more pain than pleasure at this moment. Intuition, however, the critic could contend, is against HH here: although she experiences slightly more pain than pleasure, Mary seems happy.

Note that this criticism not only applies to HH but any *theory* of happiness: any necessary and sufficient condition we postulate that happiness consists in is going to run into intuitive

difficulty with borderline cases, like this modified case of Mary, above. If just one more of Mary's desires is frustrated than satisfied, DSH deems her unhappy at the moment she is giving birth even if she thinks it is the happiest of her life; likewise with LSH and EH: if Mary is just slightly more dissatisfied with her life than she is satisfied with it or if she has just one more negative central affective state than positive ones, LSH and EH deem her unhappy in this case, too. So it is less the case that this criticism applies to HH alone than it does theories of happiness in general.

The most common response to criticisms like this is to adopt the so-called *threshold* view.⁴⁹ If HH is made into a threshold view, rather than maintaining that S is happy to the extent that she experiences pleasure, unhappy to the extent she experiences pain, it would maintain that S is happy if she reaches a certain threshold of net pleasure over pain, unhappy if she reaches a certain threshold of net pain over pleasure, and neither happy nor unhappy if her experiences of pleasure and pain fall in between the two thresholds. I do not have a problem with this modification as such; but I wonder whether it is needed. HH already recognizes that people may be neither happy nor unhappy: if S experiences the same amount, intensity, and duration of pleasure at *m* as she does pain, HH deems her neither happy nor unhappy at *m*. If she experiences slightly more pleasure than pain, HH deems her slightly happier; if S has the same number of pleasant experiences as unpleasant experiences but her pleasurable experiences are slightly more intense or slightly longer, HH again deems her slightly happier.

While HH may generate counterintuitive results on borderline cases, I doubt that it needs to be modified. Our intuitions are not so clear on borderline cases like the above Mary.

Although HH is clear on such cases, I think that our intuitions are unclear because it is difficult

⁴⁹ Haybron (2008: 147) combines a threshold view with EH and Feldman (2010: 25-26) combines a threshold view with his own hedonistic theory of happiness.

at times to determine how much pleasure you are experiencing and how much pleasure someone else is experiencing. While HH generates a clear result, it does not imply that it is easy for us to determine how happy Mary is. Now that I have addressed this criticism, let us move on to Glenda.

Before she takes the medication Glenda is displeased with many things about her life. No matter how hard she tries she cannot seem to write anything like the dissertation she wants to. Nor is she successful at acquiring the job she believes that she deserves. This makes her depressed, and her being depressed displeases her. After she takes the medication, however, her mood lifts: she goes from feeling down and out to feeling cheerful and optimistic. She was feeling lousy before taking the medicine but afterward feels good. Hence, HH deems Glenda happier after taking the medication than she was before, in accordance with intuition.

Seneca, being a Stoic, exercises control over his emotions. Thus, after delivering his rousing oration to the Senate, his mood does *not* change: he remains no more cheerful than despondent, no more tranquil than anxious. Yet Seneca does feel good: he takes pleasure over his desire to motivate the Senate being satisfied. The pleasure he receives outweighs any pain he may currently be experiencing. HH, thus, deems him happy. Lastly, in the case of Andrew, HH also deems him happy. He may receive pleasure from disgusting or immoral things; but he experiences far more pleasure than pain.

	Intuition	HH
1. Brainwashed Sally	H	H
2. Mary the mother	H	H
3. Glenda the talented	H	H
4. Seneca the Stoic	H	H
5. Andrew the swine	H	H

Unlike, DSH, LSH, or even EH, HH accords with intuition in all five cases. I contend that the case of Glenda is sufficient to make us skeptical that happiness is essentially desire-satisfaction, the case of Andrew sufficient to make us skeptical that happiness is essentially life-satisfaction, and the case of Seneca sufficient to make us skeptical that happiness essentially consists of emotions and moods. I contend that HH is a more intuitive theory than its rivals. What I need to do now is defend HH from criticism. But first I want to return to feeling and being happy.

3.1.3 Feeling and being happy

In several places since section 1.1.1 above I discussed the possible distinction between feeling and being happy. While HH does *not* imply, it certainly *permits* there to be a distinction between feeling and being happy. I think that feeling happy consists of being in a pleasurable emotional or mood state. When you are cheerful, you feel happy. This seems to accord with how we typically use the term ‘feeling happy’. But it strikes me that you may be happy without also feeling happy. When you are happy over the Seahawks winning the Super Bowl, you need not feel happy, too. While you may enjoy the Seahawks victory, you need not be in a positive emotional or mood state: your emotional or mood state could be neutral, neither positive or negative. This leaves the distinction between feeling and being happy somewhat mysterious. I am still not sure whether one may feel happy without also being happy. That said, I think that our discussion of theories of happiness, EH and HH in particular, has enabled us to get a better grasp of the possible distinction.

3.1.4 Defending HH against Haybron

In this section I explicate and respond to the criticisms of Dan Haybron. Haybron is one of HH's most recent and most able critics whose criticisms have yet to be fully considered.⁵⁰

Haybron's first objection concerns what he calls 'irrelevant pleasures', pleasures that although perhaps extremely pleasant do *not*, according to Haybron, make us happier (2008: 63).

Some pleasures, we are told, simply do not "have the slightest impact on one's happiness" (Ibid.). HH is, therefore, false, since it states that *any* pleasure S has makes her happier.

Haybron provides us with several examples of irrelevant pleasures: "[o]ne's enjoyment of eating crackers, hearing a good song, sexual intercourse, scratching an itch, solving a puzzle, [and] playing football" (Ibid.). Although orgasm may be pleasurable, likely even extremely pleasurable, this objection goes, it *need not* affect one's happiness. Furthermore, Haybron adds, "it is not just that any particular superficial pleasure seems irrelevant"; rather, superficial pleasure fails to "impact some deeper aspect of one's psychology" (2008: 63). "[N]otoriously", Haybron says, "sexual activity can leave us cold. Sometimes it just doesn't move us. This is one of the hard lessons dealt to the unsophisticated libertine, or the troubled youth seeking to relieve his melancholy through meaningless sexual encounters" (2008: 108). Intensely pleasurable experiences, like orgasm, then, may fail to move us the way central affective states like serenity and anxiety do (2008: 131).

I do not find this criticism convincing. While some pleasures, even intense ones, are brief and fleeting, Haybron discounts the possibility of brief and fleeting happiness. That is, why should we think that happiness always "impact[s] some deeper aspect of one's psychology"? Moreover, even if happiness is some deep state that is not fleeting, why not think that something

⁵⁰ Strictly speaking, Haybron does not criticize HH in particular, but rather hedonistic theories of happiness in general.

may only have a fleeting impact on one's happiness? Although HH implies that the presence of any pleasure makes you happier, the increase in happiness is proportional to several facts about that pleasure, mainly: its *intensity* and *duration*. That is, while HH implies that every episode of pleasure by itself makes you happier and that every episode of pain by itself makes you less happy, the extent to which an episode makes you happier or unhappier is determined by the episode's intensity and duration. A proponent of HH may well admit that some pleasures, though very intense, last only a very short duration and hence have only a miniscule impact on one's happiness. Such pleasures, if HH is right, *do* impact one's happiness; but the impact is only *fleeting*.⁵¹ Take orgasm: in my mind, it is often extremely pleasurable but often fleeting; it does thereby have an impact on one's happiness; but the impact is mitigated by its short duration, how quickly the pleasant feeling dissipates.

And while orgasms and other intensely pleasurable experiences may fail to move us the way serenity and anxiety do, Haybron forgets how *varied* pleasurable experiences seem. Surely, one may experience much pleasure from sensations, say, during sex but also be displeased about some state of affairs. Although the sex feels good, perhaps it does not feel as good as it used to, or perhaps you are displeased about something at work. Even though the pleasurable sensations that you have during sex contribute to your happiness, according to HH, that contribution is mitigated by anything unpleasant you are experiencing.

What Haybron needs is an example of an *intense* pleasure (pain) that is experienced over a *long* duration that does *not* appear to make us any (un)happier. I can think of such an example. Some people claim to have experienced intense pain and yet also say that the pain did not get to them, that they did not mind it. Such persons would likely deny that they were made any unhappier by their pain. I find these claims hard to believe. It seems unlikely that anyone

⁵¹ Feldman offers a similar reply to Haybron (2010: 27-28).

experiencing intense pain is happy while she is experiencing the pain. In such cases we may have good reason to believe that the pain we experienced did not make us any unhappier. Perhaps this belief better enables us to cope with the pain. Again, if Haybron is correct, we have several good reasons for thinking that we are not as in touch with how happy or unhappy we commonly think we are. So I contend that this example is just a case of someone who is not in touch with her level of happiness: even if she did not “mind the pain”, it still made her less happy.

Haybron’s second objection concerns the temporality of hedonism. Happiness, the objection goes, is forward-looking: when we say that S is happy, it entails that S is happy *now*, not necessarily that S was happy leading up to this time (2008: 69). That is, ascriptions of happiness have a certain temporality: they describe a psychological state S is now feeling; they do not necessarily describe the series of states leading up to now. If HH is correct, however, happiness is backward-looking:

to know that someone is happy on this view is only to know that his recent experience has been mostly positive. So construed, ascriptions of happiness are little more than capsule summaries or histories of subjects’ conscious episodes. They purport only to characterize the general tenor of a sequence of experiential events—namely, experiencings and liking [...] Hedonistic happiness is an essentially episodic and (in ordinary ascriptions) backward-looking phenomenon. (2008: 69)

The idea seems to be this. If we ascribed happiness to S, according to HH, we would ascribe to S a fact about her past conscious episodes, namely: that they were more pleasant than they were unpleasant. If this objection is correct happiness, however, does not have this feature; HH is, therefore, false.

But, surely, there is a distinction between *momentary* happiness and happiness over a *duration*.⁵² If HH is right, S is happy at moment *m* to the extent she experiences pleasure, unhappy to the extent she experiences pain. This is not at all backward-looking in the manner

⁵² Feldman responds to Haybron’s criticism in a similar way (2010: 31).

Haybron describes. Now, there is happiness over a duration. Suppose that duration d is composed of moments m^1 through m^3 . In order to determine how happy S is over d we need to add up the pleasure S experiences from m^1 to m^3 . Then we subtract any pain she experiences over d . So long as we have a positive result, HH deems S happy. While it is true that at m^3 we would have to look *back* at m^1 and m^2 to determine how happy S is over d , this does *not* mean that pleasure is backward-looking. Indeed, if HH is correct, ascriptions of happiness concern whether S is *currently* experiencing more pleasure than displeasure; they do not necessarily tell us anything about whether S's experiences to this moment have been pleasant.

Perhaps what Haybron means by 'backward-looking' is that pleasure is not forward-looking. Happiness is, however: it disposes us to act and be in a certain way. "If I know that Gertrude is happy", Haybron says, "then I can reasonably expect her to be a more pleasant and agreeable companion than where she otherwise disposed" (2008: 74). There is no question that happiness is sometimes forward-looking in this regard. Yet I wonder whether happiness is *always* like this. For my money, happiness is sometimes what Feldman calls 'fragile': "A person will be said to experience fragile happiness at t time iff she is happy at that time, but is also disposed to lose that happiness, or to lapse into unhappiness" (2010: 29). Think back to the example I described above in section 2.3.3. You enjoy the Seahawks' victory; but you forget about the victory on your drive home. Although you seem happy that the Seahawks won, your experience of happiness seems like an episode of *fragile* happiness, since you seem no longer happy on your drive back from watching the game. Such experiences of happiness seem common enough; if this is right, happiness is not, contra Haybron, always forward-looking.

I do think that sometimes happiness is forward-looking, that sometimes when we experience happiness, we are disposed to continue to be happy. Such happiness may only be

possible if we are in a positive central affective state, when we are in an emotional or mood state that disposes us to continue to be in that positive state. And when happiness is like this, I have no problem saying that happiness consists of the affective state; I also contend, however, that it is the *pleasure* of that affective state that more fundamentally constitutes Gertrude's happiness. Haybron, then, seems to conflate this perhaps deeper sense of happiness with happiness itself. Happiness is sometimes deep in that it penetrates our whole being in the manner Haybron describes; but it *need not do so*. And when it does do so, affective states may be the reason; but they seem to contribute to happiness only insofar as they are *pleasurable*.

Haybron's last two criticisms are his strongest; neither of them—as far as I am aware—have been debunked in the literature. We shall examine both criticisms together. They both deny that one needs pleasure (pain) in order to be (un)happy. Some *moods*, the first objection goes, affect our (un)happiness without causing pleasure (pain) (2008: 67-68). Think, for example, of irritability: when S is in an irritable mood, she thereby seems unhappy, other things equal, even if she receives no pain from being irritable. Or think of depression. “One is unhappy by virtue of being *depressed*”, Haybron tells us, “not by virtue of experiencing the unpleasantness of depression. Happiness has depth that the pleasure theory misses” (2008: 108). Some *affective states*, the second objection goes, affect our (un)happiness without causing pleasure (pain) (2008: 71). Think, for example, of sadness: when S feels sad, she thereby seems unhappy, other things equal, even if she receives no pain from the sadness.

I have two responses to this. First, I deny that S can be in an irritable mood (be sad) without also receiving some pain from being irritable (sad). Now, this is an empirical matter, a matter that is decided by the relevant psychological scholarship, scholarship which asserts that pleasure and pain are essential components of both emotions and moods. The authorities on the

psychology of emotion and moods—Barrett, Russell, Watson, and Clark—all agree on this. Take Russell for instance. He writes: “At the heart of emotion, mood, and any other emotionally charged event are states experienced as simply feeling good or bad, energized or enervated” (2003: 145). He even defines ‘mood’ as “Prolonged core affect with no Object (simple mood) or with a quasi-Object” (Ibid.: 148), where ‘core affect’ is defined as “A neurophysiological state that is consciously accessible as a simple, nonreflective feeling that is an integral blend of hedonic (pleasure–displeasure) and arousal (sleepy–activated) values” (Ibid.: 147). Or take Eisenberger who has produced research that shows that negative emotions, like the feeling of being rejected, are experienced as pain (2012a and 2012b).

In response, Haybron may well admit that affective states and moods cannot be divorced from the pleasure (pain) associated with them but nonetheless maintain that it is not the pleasure (pain) that is happiness-constitutive; rather, he may insist that it is the *affective state or mood* itself that is happiness-constitutive. If it is affective states or moods that constitute happiness, not pleasure, then it would be impossible for one to have pleasure without being in an affective state or mood and also be happy. But this strikes me as incorrect, at least when it comes to pleasure over states of affairs. Think back to Seneca. He takes pleasure over his giving a rousing oration and yet does not experience any emotions or moods. Sure, he may be happier if in addition to the pleasure he received from delivering a successful speech he also became elated and cheerful; but the fact remains: Seneca seems happier in virtue of experiencing the pleasure even though it is not accompanied by any emotion or mood.

My second response is that even if the literature is wrong and the affective states Haybron labels ‘positive’, like cheerfulness, may be experienced without their feeling pleasant, why should we think that they contribute to happiness? Refer back to Antha and suppose that

she no longer has the propensity for anxiety; suppose, rather, that she is anxious. The thing is: her anxiety is not at all unpleasant for her. Note how strange this seems. It is difficult to imagine sadness not feeling unpleasant for the person feeling sad. Indeed, if someone told us they were feeling sad but that the feeling was not at all unpleasant, we may think that he or she did not understand what ‘sadness’ means. Such a worry suggests that negative affective states, like sadness, are necessarily unpleasant. That said, it is logically possible for one to feel sad but not unpleasant: there is no contradiction between being sad and not having that sadness feel unpleasant. Intuitively, I just doubt that we are still inclined to deem Antha less happy as a result of her anxiety if her anxiety is not all unpleasant for her. Perhaps others around her who are displeased by her anxiety are made less happy because of it; but Antha’s happiness appears unaffected.

Thus concludes my discussion of Haybron’s criticisms to hedonistic theories of happiness.

3.2 Chapter conclusion

In this chapter I introduced hedonism about happiness, which postulates that you are happy to the extent you experience pleasure, unhappy to the extent you experience pain. It differs from other hedonistic theories by being neutral about the nature of pleasure. After explicating HH I showed how intuitive it is by measuring how well it accords with the five examples of happiness. Lastly, I defended it from the criticisms of Haybron. Now that we have a handle on happiness, in the next chapter I discuss its relationship with welfare.

Chapter 4: Happiness and Welfare

4.1 Introduction to welfare

The previous chapters of this dissertation concerned a certain psychological phenomenon: happiness. If I am correct, happiness just consists of pleasure and unhappiness just consists of pain. In the rest of this dissertation I am concerned with value, intrinsic value in particular. If φ is *intrinsically* valuable, φ is valuable for its own sake.⁵³ In contrast, if φ is *instrumentally* valuable, φ is valuable as a means to something that is intrinsically valuable.⁵⁴ Suppose that pleasure is the only thing that is intrinsically valuable. If this is correct, the pleasure you received from devouring that three-layered lasagna you had for lunch was *intrinsically* valuable. The lasagna was *instrumentally* valuable: it was valuable as a means for delivering you pleasure.

As I see it, there are at least three different conceptualizations of intrinsic value: aesthetic, moral, and prudential. Each value is *distinct* from the other. Take a person's life. It may be high in one value, but low in another. Suppose that Ted is a promising graduate student. He is a rather happy and successful individual who feels compelled to give something back to the world. So he joins the Peace Corps and heads off to Botswana, thereby putting grad school on hold. There he teaches about the dangers of unprotected sex. Suppose that he makes a difference in Botswana, greatly increasing the well-being of the locals. Other things equal, this duration of Ted's life is clearly high in moral value: he not only goes to Botswana with the morally right sort

⁵³ When I use 'intrinsically valuable' and 'valuable for its own sake' interchangeably throughout this dissertation, I should not be taken to deny the supervenience principle, which states that φ can be intrinsically valuable only if φ is valuable in virtue of its internal properties. I remain neutral as to whether intrinsic value entails supervenience.

⁵⁴ Not everyone agrees with this understanding of instrumental value; see Dorsey (2012a), where he claims that "An object, state, or event φ is instrumentally valuable if and only if it possesses the disposition to be instrumental to value" (146). I am inclined to think that φ is not valuable if it merely possesses the disposition to be instrumental to value; as I see it, in order for φ to be instrumentally valuable, it needs to be instrumental to value, not just be disposed to be instrumental to value.

of intention, his presence increases the overall welfare of the area. The thing is: Ted is *not* happy. He misses his home, his family, and even the intellectual rigors of graduate school. He is depressed and despondent much of the time; he just does an excellent job hiding it. In this case, it is clear that Ted does an efficient job of seeking the good of *others*; but he does not do an efficient job of seeking his *own* good. He seems to be a moral individual—at least over this duration, just not a *prudent* one. That is, Ted’s life seems low in prudential value, whereas it seems high in moral value.

It is also easy to think of lives high in prudential value without much moral value. Just think of Gordon, a Wall Street banker hell-bent on maximizing his own profit without regard for how his actions affect others. In his mind, the world be damned: so long as he gets what he wants, that is all that matters. Not only does Gordon successfully maximize his own profit, he is also esteemed highly by his peers. He has earned this esteem by corporate raiding, buying large stakes in companies only to liquidate them. Unbeknownst to the SEC Gordon relies on insider information. While he has no family, he has many friends. And while he works long hours, his nights are filled with glee and luxury. He is a very happy man. Other things being equal, such a life seems high in prudential value, low in moral value.

Here on out I am *not* concerned with value in general, but rather a particular value: *prudential* value. Prudential value differs from aesthetic and moral value in at least one crucial respect. A given piece of art—say, The Beatles’ “Let it Be”—may be more aesthetically valuable than another—say, The Beatles’ “Revolution 9”. And a person’s life—say, Ted’s—may have more moral value than another’s—say, Gordon’s. Both the Beatles’ “Let it Be” and the life of Ted may be said to possess a certain goodness. While prudential value is also a type of goodness, it is different: a person’s life that is high in prudential value is not necessarily good,

as such; rather, it is *good for* that person. If the Wall Street banker's life is prudentially valuable, it is *valuable for* him; it benefits him, makes him better off. Following the literature, I use the terms 'well-being' and 'welfare' to refer to prudential value: S has well-being to the extent S lives a life that is good for her.

Unfortunately, we have a tendency to be loose with our language when it comes to precisely what we are seeking when it comes explicating the essence of welfare. Some philosophers say that they are seeking to describe what it is for something to be intrinsically valuable.⁵⁵ This suggests that they are seeking an answer to the following question.

1) What is intrinsically valuable?

I do not, however, think that this is the question they are actually seeking. The question I think they mean to answer is the following.

2) What is intrinsically *prudentially* valuable?

These two questions are distinct. Say knowledge turns out to be intrinsically valuable, to be good, *as such*. In this case, the knowledge you possess would be intrinsically good, but it need not also be *good for you*. After all, just because something is good for its own sake does not mean that your possession of it makes you better off, at least not conceptually. An answer for (1) is not necessarily an answer for (2), and vice versa. It is logically possible for something to be intrinsically *good* without your possession of it being *good for* you. For all we know, there could be nothing of intrinsic goodness, just things that are intrinsically good for you. Or, on the other hand, there could be nothing intrinsically good for you but something that is nonetheless

⁵⁵ See, for example, Feldman (2004). He formulates default hedonism as the position that maintains "every episode of pleasure is intrinsically good; every episode of pain is intrinsically bad" (2004: 27). What he means to say is that according to default hedonism, every episode of pleasure is intrinsically *prudentially* valuable; every episode of pain is intrinsically *prudentially* bad.

intrinsically good. What I am interested in, then, is answering (2), what is intrinsically prudentially valuable.⁵⁶

I want to be clear about what my project is. I am not interested in conducting a conceptual analysis of the term ‘welfare’ in order to arrive at its definition. This is a linguistic or logical task that likely uses necessary truths that are justified a priori. Rather than engage in such an analysis, I assume that φ is good for S =df. φ benefits S.⁵⁷ I also assume that φ benefits S =df. φ makes S better off. My project is to construct a description of what it is for something to be intrinsically prudentially valuable. This is a metaphysical task, not a linguistic or logical one.

While I am trying to provide an account of the essence of welfare, I am *not* also trying to take a position on whether the proposition that expresses ‘ φ makes S better off’ is capable of being true or false. That is a second-order question that falls under the domain of metaethics, since it concerns the status of moral properties. My project here falls under the domain of normative ethics, since I try to answer the first-order, substantive, question of what makes φ instantiate a moral property.⁵⁸ Think of the utilitarian claim that φ is morally required if and only if it maximizes welfare. This is a criterion of moral rightness: it differentiates that which is morally required from that which is not. Here I am trying to discover the criterion of prudential value: that which differentiates what it is that makes a person’s life good for her from that which

⁵⁶ Dorsey (2012b) makes the same distinction: he distinguishes between relational and non-relational intrinsic value. “An object φ is relationally intrinsically good if it is intrinsically good for something or someone” (2012b: 269). “An object φ is non-relationally intrinsically good if it is intrinsically good “‘period’”, “‘tout court’” or “‘full stop’” (Ibid.). He maintains that intrinsic prudential value is relationally intrinsically valuable. If φ is intrinsically good for you, φ is relationally intrinsically good, not necessarily non-relationally intrinsically good (Ibid.).

⁵⁷ See Zimmerman (2009) for an argument as to why ‘good for’ is synonymous with ‘benefits’.

⁵⁸ I borrow this way of distinguishing between metaethics and normative ethics on the topic of welfare from Alex Sarch’s dissertation (7-8).

does not.⁵⁹ I did something similar with happiness, above. Hedonism about happiness is not necessarily a definition of the term ‘happiness’: HH does not imply that φ makes S happier =df. φ brings S pleasure. It is logically possible for happiness to have nothing to do with pleasure. So HH is not a logical or linguistic thesis about the meaning of a term; rather, it is a metaphysical account for what happiness *is*.

In this chapter I explore the relationship between happiness and welfare. I shall argue that while it seems clear that happiness benefits the happy, the benefit seems merely a pro tanto one: while I think that happiness makes the happy better off, even high levels of it need *not* make the happy well off, all things considered. I shall also argue that in order for φ to be good for you, φ need not thereby make you any happier.

4.2 The benefit of happiness

Over the summer I had the pleasure of riding in one of New York City’s subway trains. Packed in like sardines, I had but no choice to listen in on the conversation that two individuals standing next to me were having. One of them clearly seemed to be distressed. Apparently, he lost a well-paying job a few months back. Fortunately, he was able to find a new job; but there is a catch: he does not make nearly as much money as he once did. The other individual tried to assuage him, saying something like, “As long as you are happy, man, that’s what matters; everything else don’t mean squat”. We all have either heard something like this phrase or even uttered it ourselves. I wonder whether it is correct—is it true that happiness is the only thing of intrinsic prudential worth?

Well, happiness is clearly an *indicator* of prudential value. When we see an apparently happy person, we commonly conclude that that person is not only well off, but that she is well

⁵⁹ This discussion is also similar to Sarch’s (7-8).

off in rough approximation to how happy she seems. And when we see an apparently *unhappy* person, we commonly conclude that that person is not well off, in rough approximation to how *unhappy* she appears. Furthermore, '(un)happiness' is often used as a *proxy* for well-being (ill-being). In everyday conversation we commonly treat (un)happiness as a substitute for welfare (elfare). Likewise, we commonly use the term 'enjoys' ('pains') as a substitute for 'good for' ('bad for'). Someone who enjoys her wealth and success is said to be someone whose wealth and success are good for her; likewise, someone who is pained by his arthritis is someone whose arthritis is bad for him. Someone whose life is high in prudential value is often said to enjoy her life. Indeed, it is difficult to imagine a miserable person whose life is high in prudential value. That is, some minimum, *threshold*, of happiness seems necessary for well-being. It seems especially clear that some minimum threshold of happiness seems necessary for high levels of well-being.

It also seems obvious that happiness is good for you. Whenever it is present in a person's life, that person seems better off because of it. Let us think of Simon, who possesses a surfeit of both wealth and success. Let us also think of Peter, who possesses just as much wealth and success as Simon. In fact, the only difference between Peter and Simon is that Peter enjoys his wealth and success whereas Simon does not. Clearly, other things equal, Peter seems better off than Simon. Peter's happiness seems to be of some intrinsic benefit to him.

Unhappiness is similar: it seems obvious that unhappiness is bad for you. Whenever it is present, a life seems worse off because of it. Think of a person, John, who just finished leading a lecture that did not live up to his expectations. Think of another person, Paul, who also just finished leading a disappointing lecture. The difference between the two is that while Paul is unhappy about not delivering his lecture successfully, John is not: John just shrugs off the poorly

delivered lecture thinking of it as a fluke while Paul feels depressed. Clearly, John seems better off, other things equal, than Paul. Paul's unhappiness by itself seems to make him worse off.

I conclude from the above that happiness seems to be good for you, unhappiness bad for you in the sense that happiness seems at least *pro tanto* beneficial, unhappiness at least *pro tanto* harmful. What I want to do now is examine whether happiness also seems to be all-things-considered beneficial, that is, whether happiness is necessary and sufficient for welfare. To do that, I appeal to the case of Sally, discussing the nature of brainwashing in detail. If I am correct, this case shows us that happiness is *not* all-things-considered beneficial.

4.3 Brainwashing

4.3.1 The essence of brainwashing

Like the term 'happiness', 'brainwashing' is used in different ways. Sometimes we use 'brainwashing' to express mere indoctrination, when, for example, critics of Fox News accuse it of brainwashing its audience. Other times we use the term to express something more sinister, when, for example, we say that Jim Jones brainwashed his followers, causing the deaths of 909 people in Jonestown, Guyana. The sense of brainwashing that we are interested here is this latter, more sinister, one. The difference between these two senses is a difference of *kind*, not of degree: although brainwashing, as we shall see, involves indoctrination, mere indoctrination does not involve brainwashing. Some history of 'brainwashing' should be useful.

The English 'brainwashing' was coined by the journalist Edward Hunter in an article he wrote in 1950 reporting on the techniques China allegedly used to transform United States

prisoners of war into Communist sympathizers.⁶⁰ He claims to have derived ‘brainwashing’ from the Chinese ‘xǐ nǎo’, which means to wash the brain. Unfortunately, Hunter’s choice of ‘brainwashing’ may lead us to associate brainwashing with something like the ability of transforming anyone *no matter the circumstances* into a mindless slave. Such a conception implies that brainwashing is an involuntary process whereby anyone at anytime can be made into a mindless robot; this exaggerates the nature of brainwashing, making it seem unreal.

Not only may ‘brainwashing’ be misleading, the term has been used by popular culture to refer to a *recruitment* tool, a mechanism for obtaining new members. Scholarly research on brainwashing in disciplines like psychology and sociology, however, emphasizes that brainwashing is *not* a recruitment mechanism that turns anyone at anytime into a mindless slave; rather, scholars emphasize that brainwashing is a *retaining* mechanism that under the right circumstances can take some cult members and turn them into deployable agents.⁶¹ There are high costs associated with brainwashing, both to the person being brainwashed and to the person doing the brainwashing. So a cult is *not* going to want to brainwash any run-of-the-mill new recruit; rather, it is going to want to brainwash only those who have demonstrated a *commitment* that outweighs the considerable cost the brainwashing involves.⁶²

Of particular interest for us is the work of the sociologist Benjamin Zablocki. He has interviewed hundreds of ex-members and ex-leaders of cults who claim that they have been brainwashed or have brainwashed others (Zablocki 2001: 194-201). Zablocki is considered a

⁶⁰ "Brainwashing." *Propaganda and Mass Persuasion: A Historical Encyclopedia, 1500 to the Present*. Santa Barbara: ABC-CLIO, 2003. *Credo Reference*. 17 July 2008. Web. 12 Apr. 2012. <<http://www.credoreference.com.www2.lib.ku.edu:2048/entry/abcprop/brainwashing>>.

⁶¹ See, for examples, Barker and Schein (1961), Lifton (1989), and Zablocki (1997 and 2001).

⁶² It is not clear whether certain individuals are more susceptible to brainwashing than others or even whether someone may be more susceptible to brainwashing at different times of her life; rather, scholars believe “that many different kinds of people can, with enough effort, be brainwashed” (Zablocki 1998: 222).

leading scholar of brainwashing in the academy. What we are interested here is the nature of brainwashing: the essential features of what it is for a person to be brainwashed; but rather than construct a suitably general statement of necessary and sufficient conditions from scratch, we shall consider Zablocki's definition. According to Zablocki, brainwashing is:

a set of transactions between a charismatically-led collectivity and an isolated member of the collectivity with the goal of transforming the member into a deployable agent [...] [where a] deployable agent is one who can be relied on to continue to carry out the wishes of the collectivity even when they are in opposition to his or her own simple hedonic interests and in the absence of any external controls. (2001: 239)

Zablocki's understanding of 'deployable agent' aptly describes the submission involved when one person is brainwashed by another. It is worth quoting Zablocki again in full.

A deployable agent is one who evaluates his self-interest rationally as the group would wish. It does not argue the elimination of choice but rather the modification of the preference structure on which choice is based. The brainwashed individual remains capable of rational choice and action but over a transformed substrate of values and preferences remapped to conform to the collective ideology. (1997: 97)

Notice that brainwashing transforms the values and preferences of individuals to conform to the group.⁶³ The brainwashed equate their self-interest with the group's interest. So the brainwashed are capable of making rational choices in the sense that they promote the group's interest. Zablocki's use of 'charismatic' is also insightful: the authority figure must be perceived by the person being brainwashed as trustworthy and even capable of performing extraordinary deeds (2001: 241). Nonetheless, Zablocki's definition has two problems.

First, although most brainwashing occurs in groups, it is possible for one individual to brainwash another without operating in a group. Some therapists are alleged to have brainwashed their patients; if Zablocki is correct, we would have to apply a different term to such cases *merely* because the setting is not a group one; but this is an insufficient reason.

⁶³ Even one's deepest values and beliefs may be changed as a result of the brainwashing: "The core hypothesis is that, under certain circumstances, an individual can be subject to persuasive influences so overwhelming that they actually restructure one's core beliefs and world-view and profoundly modify one's self-conception" (Zablocki 1997: 99).

Second, his ‘with the goal of’ should be dropped: brainwashing is just a process, *not* necessarily an intentional process. The degree to which the process occurs, not the degree to which the process is intended to occur, is the degree to which brainwashing occurs.

Zablocki writes that brainwashing may occur even when it opposes the “simple hedonic interests” of the brainwashed. At first glance, this may seem to conflict with brainwashed pleasure: if brainwashing happens contrary to a person’s hedonic interest, how can one find brainwashing pleasurable? This misconstrues Zablocki’s definition: he is *not* saying that all brainwashing occurs contrary to a person’s hedonic interests; he merely says that brainwashing *can* occur contrary to a person’s hedonic interests. And this is compatible with brainwashed pleasure: although brainwashing may occur such that the brainwashed receive little to no pleasure from it, it is also possible for a person to be brainwashed and receive pleasure from it. Zablocki’s ‘set of transactions’ highlights this: there is something that the person receives in return from being brainwashed.

Considering the above, Zablocki’s definition should be modified such that ‘brainwashing’ is defined as: *the set of transactions by which a charismatic authority transforms another person into her deployable agent*. This, then, is the essence of brainwashing. Now that we have a theory let us consider the process by which brainwashing operates.

4.3.2 Brainwashed Sally

Refer back to Sally, who, like most who are brainwashed, is a young adult, in her second year of college. She joins a parachurch organization at her college. At one of the organization’s events she meets Dianne, whose church Sally’s friends attend. Something about Dianne just seems to click with Sally, leading them to share several intense conversations. Soon enough Sally begins attending Dianne’s church.

Now, this is a small church, consisting of around just twenty members. In fact, it just meets in Dianne's home on Sundays. While the smallness of the church initially concerns Sally, the positive feelings she receives from attending it overwhelm her worry. It does not take long for Dianne to realize that Sally would be an asset to the church, someone to help it attract new members. So Dianne begins brainwashing Sally fairly early. According to Zablocki, brainwashing involves an intentional systematic psychological assault consisting of repeated cycles of affirmation and denial, confession and reconfession, a non-linear process of "intermediate steps and backsliding" (2001: 238). This assault has three phases: 1) stripping away of Sally's beliefs present before the brainwashing, 2) Sally identifying and mimicking the behavior of others under Dianne's authority,⁶⁴ and lastly 3) the death of the old Sally along with the birth of a new Sally when she has internalized her new beliefs (2001: 223).

In phase (1), Dianne indoctrinates Sally, getting Sally to believe what she says by appealing to her emotions rather than her critical faculties.⁶⁵ Through a series of Bible studies that weave together enough passages in just the right ways Dianne gets Sally to believe that she (Dianne) has the only correct understanding of the scriptures. When Sally suggests alternative interpretations, she is scolded. And when Sally fails to ask critical questions, Dianne rewards her by emotional affirmation. This affirmation of course feels great to Sally. She quickly learns to do what pleases Dianne and avoid doing what displeases her. Following Dianne's lead, the other members of the church also affirm Sally. Soon enough, Dianne is cutting Sally off from the

⁶⁴ As phase is two is phrased, it may seem that Sally can only be brainwashed by Dianne in a group setting; but this is not the case: if Dianne is brainwashing Sally and no others, this phase would instead consist in Sally identifying and mimicking the behavior Dianne desires.

⁶⁵ Before the nineteenth century, both English verbs 'indoctrinate' and 'doctrinate' meant merely to teach or instruct ("indoctrinate, v." OED Online. September 2012. Oxford University Press. 6 October 2012 <http://www.oed.com.www2.lib.ku.edu:2048/view/Entry/94678>). Starting in the early 1800's, however, 'indoctrinate' was used in pejoratively to mean teaching uncritically, getting S to believe something without appealing to S's critical capacities.

outside world. Where Dianne (and by extension, her church) has the correct understanding of the Bible, the rest of the world's understanding is incorrect. Since church members have the only correct understanding of the Bible, Sally has no need to refer to outsiders. The first few times she does refer to outsiders, she is scolded; so that soon stops. Divorcing Sally from outside sources of information enables Dianne to obtain Sally's trust, which Dianne takes advantage of to begin removing Sally's beliefs present before her joining the church. Her conversations with Dianne now do not merely concern the scriptures but also concern other areas, say, her boyfriend who she was then dating for over three years. This moves us to phase (2).

Phase (2) not only consists of Dianne changing the content of Sally's beliefs, it also consists of Dianne changing the credence Sally assigns to her beliefs relative to one another. Other things equal, changes in a person's credence level is not necessarily bad for that individual; other things, however, are *not* equal: Sally's assigning new credence levels is the sole result of Dianne's manipulations. Dianne only accomplishes the change by dulling Sally's critical faculties, using logical fallacies and emotional ploys to appeal to Sally's affective capacities.

Now, Sally also suffers over this period: she experiences at times intense grief and anxiety. After all, she comes to believe that her family and friends outside of the cult are not saved. Dianne also has Sally break up her relationship with her boyfriend. Sally comes to regret her previous life, the one spent pursuing her own projects rather than God's. Yet, all things considered, the love and affection Sally receives from Dianne and the rest of the cult far outweighs any suffering she experiences. This leads us to phase (3).

Sally's new beliefs and credence levels change her identity: how Sally essentially regards herself changes as a result of the brainwashing. This involves a power imbalance: if Dianne

brainwashes Sally, Dianne has power over Sally such that Dianne has influence over every aspect of Sally's life, leaving Sally without privacy with regard to Dianne. By the end of phase (3), Dianne becomes Sally's sole source of information, leaving Sally no longer able to cognitively or emotionally deindividuate herself from Dianne. The result of this nine-month long process leaves Sally a mere deployable agent of Dianne, willing to do anything Dianne asks of her. She is willing to drop out of college, abandon her degree, and spend the rest of her life spreading Dianne's gospel. Sally's deepest beliefs and values have changed. Her parents and those who were close to Sally in the past would say that she is no longer the same person, that *who* she is has changed.

4.3.3 The harm of brainwashing

Now that we have discussed the essence of brainwashing and how it operates, it should be clear that it seems bad for the person being brainwashed. Unlike 'happiness', which is ambiguous between psychological happiness and well-being, of which only the latter is evaluative, 'brainwashing', then, is an evaluative term. It just so happens that rather than connote positive value, like 'well-being', 'brainwashing' connotes negative value: if Dianne is brainwashing Sally, Sally is being transformed into a mere deployable agent of Dianne's. This process harms Sally: Sally is better off, other things equal, not being transformed into Dianne's mere deployable agent. This is because brainwashing violates the brainwashed: it penetrating deep into their psyches, changing their identities. It takes a person who was once capable of thinking independently, forming her *own* beliefs and desires, and then transforms her into someone who cannot think for herself but instead performs whatever it is that the persons doing the brainwashing ask of her.

In Sally's case, the harm that the brainwashing does to her, we commonly believe, *justifies* her parents' intervening: they intervene on *Sally's behalf*. Sally's parents hire counselors to help their daughter consider whether Dianne has absolute control over her. During the next few weeks Sally comes to realize that she has been brainwashed. Fortunately, then, the intervention worked in freeing Sally from Dianne's control. We typically deem such interventions justified on the grounds that the brainwashed are better off not being brainwashed. If Sally was not better off years after the brainwashing but was instead better off brainwashed, with what right do her parents have in interfering with Sally's involvement with the group?

4.4 Hedonism about happiness and welfare (HHW)

Now that I have described what brainwashing is, I shall consider whether happiness is all-things-considered beneficial. In chapter three I argued that hedonism about happiness (HH) was true. If this correct, happiness just is pleasure. So to determine whether happiness is all-things-considered beneficial, let us consider hedonism about happiness and welfare (HHW), which consists of two claims: 1) you are happy to the extent that you have pleasure, unhappy to the extent you have pain⁶⁶ and 2) you are well off (worse off) to the extent that you are (un)happy. HHW is a strong claim: it says that in all cases, prudential value is determined only by pleasure; the more of it that you have, the better off you are, period. If this is correct, the pleasure Sally has while brainwashed makes her all-things-considered well off.

I want to be precise about what I mean by 'all things considered' and 'pro tanto'. Φ makes S pro tanto better off if and only if, other things equal, ϕ makes S better-off. Φ makes S all-things-considered better off if and only if, taking everything else into account, including

⁶⁶ HHW, like HH, does *not* take a position on the essence of pleasure: it says that whatever pleasure essentially turns out to be, that is the essence of happiness, which in turn is the essence of welfare.

whatever is not equal, φ makes S better off. Notice that if φ makes S all-things-considered better off, φ also makes S pro tanto better-off; but if φ makes S *merely* pro tanto better off, φ does *not* make S all-things-considered better off.

Let us return to brainwashed Sally, who is thoroughly *pleased*. Never has she *felt* so spiritually uplifted, so alive with power, so at ease with people, so confident of the truth of her convictions, and just so *fulfilled*. Not only does she feel great: brainwashed Sally also takes pleasure in feeling so great. Indeed, she feels so pleasant and takes so much pleasure in feeling so pleasant that Sally has a higher surplus of pleasure than pain, the highest surplus she has ever had over a similar duration.

Also remember that Sally's family conducts an intervention on Sally's behalf. Her intervention goes well, so well that Sally leaves the cult. Over the next several months Sally struggles to cope with the loss of leaving the cult. Sally is often displeased: she feels spiritually deflated, alienated from others, unsure of what convictions to hold, let alone be confident of, and just so unfulfilled. She realizes that she was manipulated by the cult and hence feels anxiety over not knowing whom she can now trust. This leaves Sally feeling downright miserable, so miserable that she even contemplates suicide from time to time. Not only does she often feel lousy: the fact that she feels lousy is displeasing to her, too. Sally begins to see a therapist. With the therapist's help she begins to do something she could not do while brainwashed: exercise her own critical capacities, reflecting over the life she wants to lead, considering what evidence she has for her beliefs, and so on. Exercising her capacities in this manner feels good to Sally, and she takes pleasure in once again relying on her own thinking.

Fast-forward several years: Sally is still out of the cult. She no longer has difficulty coping with it, or at least not nearly to the extent she once did. More often than not Sally is

pleased: she feels content, sure of her convictions, and just fulfilled. Not only does she feel good, the state of affairs of her feeling so good brings Sally pleasure, too. She also continues to exercise her critical capacities; indeed, she is now a better critical thinker than she was previous to her brainwashing. Sure, there are things she wishes she had done otherwise; but for the most part, she is content. Indeed, if given the choice to continue the brainwashed-free life she now leads or go back to the brainwashed life, Sally would choose to continue her current life. That said, she did feel better and was more satisfied while brainwashed than she is now: while she does have much more pleasure now than pain, Sally just had even more pleasure while brainwashed. Sure, Sally experienced more pain while being brainwashed than she does now, but the increase in pleasure she experiences outweighs the pleasure she experiences such that the net amount of pleasure she experienced while brainwashed is significantly higher than what it is now.

Focus on the three periods of her life: the year of her brainwashing, the first several months after she leaves the brainwashing, and several years after she has left the brainwashing. Of the three periods, she has the most overall pleasure while brainwashed. If HHW is correct, pleasure is not only the essence of happiness, it is also the essence of welfare. Thus, of the three periods of her life, HHW deems Sally best off brainwashed. But does this seem right? Well, Sally's brainwashed happiness does seem to provide some benefit to her. Setting aside all we know about brainwashing, if we just look at the pleasure Sally receives, it seems to make her better off. Suppose that there is some world where Sally is brainwashed but does not receive the pleasure she does in ours. Other things equal, our Sally is better off: at least our Sally receives pleasure from being brainwashed; the other Sally receives none at all.

However, given what we know about the nature of brainwashing from section 4.3, above, even though Sally's pleasure seems beneficial to her, the benefit seems outweighed by the harm that the brainwashing does to her. Where she was an excellent student bent on going to medical school, she is now the pawn of Dianne. Where she was exercising her natural talents autonomously, she now does whatever it is that Dianne tells her to do.

Now the proponent of HHW may have a ready retort: she could claim that we typically think that brainwashing is bad for the brainwashed because it causes more pain than pleasure, at least in the long term. Consequently, while our aversion to brainwashing is justified, we mistakenly think it is intrinsically bad for us, when it is the *pain* that typically follows from the brainwashing that is intrinsically bad for us. In the rare event that the brainwashed feel more pleasure than pain, despite our thinking otherwise, the brainwashed are actually well off, all things considered.

I doubt that this is true, however. That is, I doubt that brainwashing tends to bring more pleasure than pain to those undergoing the brainwashing. Sure, it brings more pain than pleasure to the friends and family members of the brainwashed: they struggle with the loss of a loved one to an ideology. But the brainwashed receive a benefit from the brainwashing; why else would they undergo it? This benefit consists of the feelings of affirmation and affection that the cult provides. The brainwashed also benefit from a living a narrative that explains precisely why they exist and what their purpose is. Such a narrative feels electrifying: instead of being burdened by meddlesome critical questions, you can just rest in the peace that surpasses all understanding.

The problem that HHW faces is that while pleasure seems to make us better off, some pleasures seem to make us better off than others. The case of Sally emphasizes this: all things

considered, Sally seems better off enjoying the fruits of, say, being that medical student she so desperately wanted to be prior to the cult, than she does receiving pleasure from being brainwashed. HHW cannot account for this: it says that *any* pleasure you experience is just as good for you as any other, so long as they are of the same intensity and duration.

In response to this criticism, a proponent of HWW could modify her theory. She could acknowledge that Sally is better off years after her brainwashing than she is while being brainwashed even though she has more pleasure brainwashed. The theorist could maintain that there is something *defective* about the brainwashed pleasure that makes it inferior to non-brainwashed pleasure. The most popular way of going about such an account is to insist that there is a distinction between *authentic* happiness and *inauthentic* happiness. This is where we shall turn.

4.5 Authentic happiness theories of welfare

To start, there are two ways to formulate an authentic happiness theory of welfare. It could be formulated such that *only* authentic happiness is intrinsically good for you. Formulated in such a way authentic happiness would be necessary and sufficient for well-being. If this were true, inauthentic happiness would have no intrinsic prudential value; but this seems false.

Remember, Sally seems better off brainwashed and happy, other things equal, than brainwashed and unhappy. Thus, it seems like the preferred way to formulate an authentic happiness theory of welfare is by maintaining that *all* happiness is intrinsically good for you; it just so happens that *authentic* happiness is better for you than *inauthentic* happiness. This is because the extent to which happiness makes you better off is determined, at least in part, by how authentic it is.

4.5.1 Authentic hedonism about happiness and welfare

Proponents of *authentic* HHW need to formulate what it is for some pleasurable experiences to be authentic and then give more weight to these experiences than inauthentic ones. Luckily for the hedonist, Fred Feldman provides two such accounts: 1) truth-adjusted attitudinal hedonism and 2) desert-adjusted attitudinal hedonism.⁶⁷ Notice that both employ the term ‘attitudinal’. This is because both presuppose an *externalist* understanding of pleasure. That is, both imply that pleasure is essentially an attitude, not a feeling or sensation. I shall explicate the former theory first, then move on to the latter.

4.5.1.1 Truth-adjusted HHW

Truth-adjusted attitudinal hedonism about welfare (‘truth-adjusted HHW’, for short) maintains that while pleasure is the essence of welfare, to what extent a particular episode of pleasure contributes to an individual’s welfare is determined by the pleasure’s intensity, duration, and whether it is *based on true propositions*. Pleasures based on true propositions are more intrinsically prudentially valuable than pleasures based on false propositions. Interestingly enough, truth-adjusted HHW makes no such claim about displeasure: it maintains that displeasure is bad for the individual, no matter if it is based on truth or falsehood.⁶⁸ According to truth-adjusted HHW, then, S is happy to the extent that she experiences pleasure; S’s life goes well for her to the extent she is happy; and the extent to which a particular episode of happiness makes her better off depends, at least in part, by whether it is based on truth.

⁶⁷ Feldman formulates both what he calls ‘truth-adjusted intrinsic attitudinal hedonism’ and ‘desert-adjusted intrinsic attitudinal hedonism’, both of which he first links to welfare, not happiness (2004). He later offers a theory of welfare and happiness that maintains happiness is intrinsic attitudinal pleasure and that welfare is desert-adjusted happiness (2010: 210-215).

⁶⁸ I have opted to use ‘displeasure’ here, not ‘pain’. I used ‘pain’ in chapter three at least in part because I did not want to favor internalism or externalism. But since truth-adjusted AHW implies that pleasure is attitudinal, there is no longer a reason from me to use ‘pain’.

I think the following example will help us understand what is for an episode of pleasure to be based on truth. Imagine two individuals, Bruce and Ted: both are blissfully happy and successful businessmen who love their wives and children. The *only* difference between the two is that Ted's happiness is based on a lie: his wife is cheating on him. Intuitively, it seems that Bruce's life is going better for him than Ted's. Yet hedonism about happiness disagrees: since both are equally happy, it deems both *equally* well off. Truth-adjusted HHW is different, though. Because Ted's wife is cheating on him, his pleasure from his thinking that his wife is loyal is based on false beliefs. Bruce's pleasure, however, is based on true beliefs. And since it is based on true beliefs, truth-adjusted HHW deems his pleasure more prudentially valuable and therefore deems him better off (than Ted). While this theory accords with intuition in the case of Bruce and Tim, let us see how well it does with Sally.

Brainwashed Sally sure does believe many falsehoods. She believes that Dianne is just about infallible, that she can do no wrong. Sally also believes that she is flourishing and helping others to flourish, as well. And, of course, she receives pleasure from all these false beliefs. Basically, Sally is blissfully unaware of the fact that she was manipulated into joining a destructive cult. That said, much of her pleasure is based on *truth*. For example, she takes much pleasure over the fact that she is a leader of her church, in fact, the right-hand gal of Dianne. For another, the fact that she recruits new members on a weekly basis also brings Sally pleasure. If the majority of Sally's (intense and long-lasting) pleasure were based on false beliefs, truth-adjusted HHW would deem that pleasure less prudentially valuable than if it were based on true beliefs. And it would, therefore, deem Sally better off basing her pleasure on true beliefs, that is, it would deem her better off not brainwashed. But the majority of Sally's (intense and long-lasting) pleasure is based on *true* beliefs. So truth-adjusted HHW is not of much help to the

hedonist: brainwashed Sally's truth-adjusted happiness is greater than Sally's truth-adjusted happiness years after the brainwashing, meaning that truth-adjusted HHW deems Sally better off brainwashed.

The proponent of truth-adjusted HHW could respond with something like the following. Surely Sally's pleasure is *ultimately* based on falsehood. She believes that she is not being brainwashed, or at least that her discipleship with Dianne is not bad for her. And she would not believe that she is a powerful leader or recruitment machine without also believing that her brainwashing is good for her. So *at bottom* her pleasures are based on false beliefs!

I admit that in some ultimate sense Sally's pleasure is probably based on falsehood; but I fail to see how this is relevant. Look, Sally's believing that she is a powerful leader and recruiting machine—both beliefs of which she takes pleasure over—are causally related to false beliefs in the sense that she would not have those beliefs without first having her false ones; but so are nearly all of our beliefs! Most—if not all—of our beliefs are causally related to false ones. Hardly any pleasure appears to be *completely* based on true beliefs. So truth-adjusted HHW faces a dilemma: either it deems brainwashed Sally well off, all things considered, or it implies that hardly any of our pleasure is good for us. Either way, truth-adjusted HHW seems false.

There is an additional famous case that truth-adjusted HHW fails to accord with intuition on, a case we have already discussed: Andrew the swine. He has tons of pleasure. He enjoys using illegal drugs, fornicating with the hired hands, and bestiality. The thing is: these pleasures are all clearly based on facts, not falsehoods. The problem with these pleasures does not appear to be that they are based on falsehoods, since they are not; rather, the problem seems to be something else. Allegedly, the pleasures are just too shallow, they are beneath Andrew. Sure,

the pleasure seems intrinsically good for him; but Andrew seems even better off getting pleasure from more worthwhile things, from things of *higher desert*. This brings us to desert-adjusted HHW.

4.5.1.2 Desert-adjusted HHW

Desert-adjusted attitudinal hedonism about welfare ('desert-adjusted HHW', for short) maintains that while pleasure is the essence of welfare, the extent to which a particular episode of pleasure contributes to an individual's welfare is determined by the pleasure's intensity, duration, and the *worthiness* of the object of pleasure: pleasures based on objects of high desert are more prudentially valuable than pleasures based on objects of low desert. That is, the prudential value of a pleasure taken from something good or beautiful is increased; and the prudential value of a pleasure taken from something bad or ugly is mitigated. Likewise, displeasures based on objects of low desert are less harmful than displeasures based on objects of high desert. That is, the harm of a displeasure taken from something bad or ugly is mitigated; and the harm of a displeasure taken from something good or ugly is increased. According to desert-adjusted HHW, then, S is happy to the extent that she experiences pleasure; S's life goes well for her to the extent that she is happy; and the extent to which a particular episode of happiness makes her better off depends, at least in part, by how deserving the object of S's pleasure is.

Refer back to Andrew. Sure, he has a lot of pleasure; but the pleasure is over disgusting things: illegal drugs and bestiality. These objects are of *low* desert. Hence, while desert-adjusted HHW deems the pleasure arising from them prudentially valuable, the value is mitigated by the low worth of the illegal drugs and bestiality. Thus, while all pleasure that Andrew experiences makes him pro tanto better off, only pleasure based on things of high desert make him all-things-

considered well off. So according to desert-adjusted HHW, while Andrew is better off experiencing the pleasure he does than none at all, he is even better off taking pleasure in objects of high desert. This accords well with intuition. How does desert-adjusted HHW fare with Sally?

Again, I think that the hedonist is in trouble. Much of what Sally takes (intense and long-lasting) pleasure in is good, or at least *not* obviously bad or ugly. She takes pleasure over being a part of something bigger than herself, of feeling that her life is infused with purpose and meaning, of recruiting new members to experience what she experiences, and of being a leader and having others look up to her. Sure, these objects may be bad in the sense that they just further Sally's own involvement in the cult and get new members to join it. But this at best gets us *instrumental* badness or ugliness: the objects are not obviously *intrinsically* bad or ugly; the objects are just means to badness or ugliness. The prudential value of brainwashed Sally's desert-adjusted pleasure, then, outweighs the prudential value of her desert-adjusted pleasure years after she leaves the brainwashing. Hence, desert-adjusted AHW deems Sally not only pro tanto better off brainwashed, but also all-things-considered well off brainwashed, in conflict with intuition.

In addition to conflicting with intuition in the case of Sally, I wonder whether desert-adjusted HHW conflates prudential value with something like moral value. Some immoral and ugly things seem to make us better off. Think back to Gordon, the Wall Street banker hell-bent on maximizing his own profit, without care for how doing so affects others. His life seems to be high in prudential value, low in moral value. Gordon's corporate raiding and insider trading give him success but bring suffering to many others. He seems to lead a successful life while he engages in immoral and ugly activities. While Gordon's actions seem to hurt the good of others,

they seem to further his own good just as much as engaging in moral or projects would. I find it difficult to see how a proponent of desert-adjusted HHW can go without appealing to moral value in describing what she means by ‘bad’ and ‘ugly’.

Both of the popular formulations, then, that a hedonist may make in order to capture the idea of authentic pleasure seem unsuccessful: both truth-adjusted and desert-adjusted HHW deem Sally well off brainwashed, all things considered.

4.6 Chapter conclusion

In this chapter we explored the relationship between happiness and welfare. Happiness clearly seems to be an indicator of welfare. Moreover, ‘happiness’ is often used as a convenient proxy for welfare. Happiness just seems good for the happy. *But* the benefit happiness provides seems to be merely a pro tanto one. I argued for this by appealing to the nature of brainwashing and the case of Sally. While Sally seems happiest brainwashed, she does not seem well off, all things considered. I then introduced the notion of *authentic* happiness and discussed how HHW could be modified to incorporate it in two different ways. But both modifications failed to make HHW accord with intuition in the case of Sally. Hedonism, then, although right about happiness, seems to miss the mark on welfare: it describes the accurate psychological state that is the essence of happiness, but it implies that that psychological state is, seemingly incorrectly, all-things-considered beneficial to the happy. Thus, while happiness seems intrinsically good for us, it does not seem to be the only thing intrinsically good for us. If it were, the benefit that Sally’s happiness brings her could not be outweighed by the harm of her brainwashing. But because she does not appear to be well off, all things considered, happiness does not appear to be the only intrinsic prudential good.

Chapter 5: What Welfare is Not

In the above chapter I criticized hedonism about happiness and welfare. In this chapter I continue explicating and criticizing theories of welfare; I focus on five of the more popular theories. Each of the following five either fails to accord with intuition in the case of Sally or has other problems. In the next chapter I shall explicate and defend my own theory of welfare.

5.1 Desire-satisfaction about welfare (DSW)

5.1.1 The essence of DSW

DSW is the position that the essence of welfare is the satisfaction of your desires: you are well off to the extent that your desires are satisfied, worse off to the extent they are frustrated. If DSW is correct, the satisfaction of *any* of your desires benefits you, period. Proponents of DSW use ‘desire’ and ‘satisfaction’ in much the same way that proponents of DSH use them; that said, we should review both.

Desires are pro-attitudes: to desire ϕ is to be positively disposed to ϕ . While you may be aware of a desire you have, you need not be. You seem to desire many different things *without* being vividly aware that you do, namely: breathing, hunger, thirst, and so on. DSW maintains that the extent to which a particular satisfied desire contributes to your welfare is determined by the *intensity* and *duration* of the desire: basically, the more intensely and the longer you desire ϕ , the more satisfying your desire will increase your well-being. There is propositional content to a desire: whenever you desire something, you want the content of your desire to obtain.

Now that we have reviewed desire, let us move on to satisfaction. While there were two different ways to understand satisfaction under DSH, there is just one way to understand it under

DSW: a desire for ϕ is satisfied if and only if ϕ . Thus construed, DSW maintains that you are well off to the extent that you get what you want, worse off to the extent that you do not.

5.1.2 The *pro tanto* benefit of satisfied desires

Like happiness, I think that satisfied desires are an *indicator* of welfare. When we encounter a person who seems to have satisfied many of her desires, we commonly conclude that that person is well off; and when we encounter a person who seems to have many of his desires go unsatisfied, we commonly conclude that that person is not well off. We also commonly define success in terms of getting what we want; and we commonly use ‘success’ as a convenient *proxy* for welfare. Further, it is difficult to imagine a person who fails to get much of what she wants whose life is high in prudential value. That is, some minimum, *threshold*, of satisfied desires seems necessary for well-being. And it seems clear that some minimum threshold of satisfied desires seems necessary for high levels of well-being.

It also seems obvious that satisfied desires are intrinsically good for you. Whenever a person gets what she wants, that person thereby seems better off. Think back to Simon and Peter who possess a surfeit of both wealth and success (section 4.2, above). The only difference between the two is that Peter *longed* for his wealth and success whereas Simon did not. Other things equal, Peter seems better off than Simon: Peter’s satisfied desires seem to be of some benefit to him.

While satisfied desires do seem to be of some benefit to you, they do *not* seem to be beneficial, all things considered. Refer back to Sally.

5.1.3 DSW and Sally

There are three relevant periods of Sally’s life: her brainwashing, her first couple of months after leaving the brainwashing, and several years thereafter. Of the three periods Sally

clearly seems happiest while brainwashed, when she experiences the most (net) pleasure. Yet Sally seems better off years after the brainwashing, when even though she is less happy, she relies on her own rationality and critical thinking. Of the three periods, Sally seems least well off the first couple of months after her brainwashing when she is depressed and downright miserable.

She not only receives much pleasure while brainwashed, she also satisfies many of her desires, so many that she has an abundance of satisfied to unsatisfied desires. In the first month or two after she leaves the cult, although Sally has more desires satisfied than unsatisfied, she has many more satisfied desires while brainwashed. And even though Sally has more net desire-satisfaction several years after she leaves the brainwashing than in the first month or two after leaving the cult, she still has much more net desire-satisfaction while brainwashed. If DSW is correct, the satisfied desires Sally has while brainwashed make her all-things-considered well off.

Now, I admit that Sally's satisfied desires are of some benefit to Sally; but the benefit, I maintain, is *merely pro tanto*: once we consider what brainwashing is, the benefit that the satisfied desires provide Sally with seems outweighed by the harm that the brainwashing does to her. Even though she has more (net) desire-satisfaction while she is being brainwashed than several years after the brainwashing ends, Sally seems better off several years after.

The problem that DSW faces is that while satisfied desires do seem to make us better off, some satisfied desires seem to make us better off than others (independent of their intensity and duration). Other things equal, Sally seems better off satisfying the desires that stem from her independence, less well off satisfying the desires that stem from her brainwashing.

The proponent of DSW has a response to this criticism. Refer back to HHW, the position that well-being just consists of happiness. It, like DSW, failed to accord with intuition in the case of Sally. So I modified it by distinguishing *authentic* from *inauthentic* happiness, where the former is more intrinsically prudentially valuable than the latter. We can make a similar move with DSW.

5.1.4 Full-information DSW

A desire-satisfaction theory of welfare that makes the distinction between authentic and inauthentic desires distinguishes between a person's current wants and her deepest wants. Deep desires are supposed to be something like S's true desires, desires S adopts of her own accord, not desires S adopts due to external influences. S's deepest desires, then, are meant to be her *authentic* desires, desires that come from who she *really* is. If we can plausibly explain that Sally's brainwashed desires are merely her current desires, not her deepest ones, we might be able to plausibly say that she is better off not brainwashed.

The most plausible formulation of an *authentic* desire-satisfaction theory found in the literature relies on the notion of full information. The chief proponents of a full-information desire-satisfaction theory about welfare, Richard Brandt and Peter Railton, maintain that S is well off to the extent that she satisfies desires that a *fully informed* and *fully rational* S would want the actual S to want (for their own sake) (Railton 1986: 17).⁶⁹ We need to discuss how proponents of this view understand both 'full information' and 'full rationality'.

⁶⁹ Railton contends that "an individual's good consists in what he would want himself to want or pursue, were he to contemplate his situation from a standpoint fully and vividly informed about his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" (Railton 1986: 17). Railton also says that "an individual's *intrinsic* good consists in attainment of what he would in idealized circumstances want to want for its own sake—or, more accurately, to pursue (for wanting is only one way of pursuing)—were he to assume the place of his actual self" (Railton 1986: 17-18).

S is informed to the extent that she is aware of the effects that her desires have on her own life; S is fully informed if and only if she is made fully aware of such effects. Full information is intended to capture what Connie Rosati calls the ‘birds-eye’ perspective: a perspective from which we stand back and assess the different lives generated by satisfying our desires (Rosati 1995: 297). Once fully informed we then choose to satisfy desires that lead to the most appealing life. To be fully rational is to be “entirely free from cognitive error or lapses of instrumental rationality” (Railton 1986: 16). A person follows the commands of instrumental rationality to the extent that she does what she has reason, not necessarily overriding reason, mind you, to do (Railton 1986: 7).⁷⁰ Cognitive errors consist of, for example, presentation bias, when the ordering of the information alone influences understanding it (Railton 1986: 21-22). According to full-information DSW, then, S is well of to the extent that she satisfies her fully informed desires.

I earlier argued that although Sally has more net desire-satisfaction brainwashed than she has either months or several years after the brainwashing, she seems better off several years after she left the brainwashing behind. If this is correct, DSW is false. We modified DSW to full-information DSW in order to accommodate the case of brainwashing; does the modification work?

For brainwashed Sally to satisfy her most intense desires, she must remain in the cult. Now, a fully informed and fully rational Sally could step back and assess (entirely free from cognitive errors and lapses of instrumental rationality) the myriad of lives generated by satisfying certain desires. If Sally were to become fully rational, she would no longer be fooled

⁷⁰ Railton has two reasons for thinking that full rationality does not necessarily consist of overriding reason. One, requiring S to perform only those acts that she has *overriding reason* to perform presupposes that there is some uncontroversial “logic of induction” that assigns univocal degrees of epistemic warrant to any proposition (1986: 6); the existence of such a logic is dubitable. Two, even if such a logic exists, he maintains, the mere fact that S is justified in believing *p* does *not* imply that S should always obey *p* (Ibid.).

by Dianne's tricks. If she were to become fully informed, Sally would become aware of the consequences of her possible actions. At the precise moment she becomes fully rational and fully informed, then, Sally also becomes aware that she is being brainwashed. Once she becomes aware of all of the consequences of her actions, she sees that the period of her life thus far in the cult has been the happiest of her life. And she understands that she would be happiest if she continues her life inside rather than outside the cult. But Sally does not value her happiness above all else; there are things she values more, like: living her life apart from anybody's control, doing what she wants to do. So a fully informed and fully rational Sally desires brainwashed Sally to leave, not stay part of, the group. According to full-information DSW, Sally is well off to the extent that she satisfies the desires that a fully informed and fully rational Sally would want brainwashed Sally to satisfy. A fully informed and fully rational Sally would want Sally to live free of the brainwashing; this requires Sally to leave the cult. Thus, a fully informed and fully rational Sally would want Sally to leave the cult. Full-information DSW, then, deems Sally better off years after the brainwashing than during it, in accordance with intuition.

5.1.5 Criticizing full-information DSW

While it is good that full-information DSW accords with intuition in the case of Sally, something seems amiss. With full information, Sally understands both that she is being brainwashed and also that she needs to continue being brainwashed in order to maximize her happiness over her lifetime. It so happens that while she desires to be happy, she does *not* desire it above all else.

But what if Sally were different?—what if she valued her happiness above all else? Let us imagine Molly, who is identical to Sally who finds herself in identical circumstances as Sally except that if she were to become fully informed and fully rational at any time during her

brainwashing, she would choose to remain in the cult: Molly, unlike Sally, values her own happiness above all else. When she reflects about the most pleasant periods of her life, Molly decides that she wants to bring about as many of those experiences as she can, even if it takes being brainwashed to do so.

What does intuition say about Molly? There is a difference between Molly and the non-fully-informed Sally: Molly makes her decision free of the influence of Dianne. In an important sense, Sally does *not* choose to become Dianne's mere deployable agent; Molly does *choose*, however. That said, I think that we are still pretheoretically inclined to judge Molly better off years after the brainwashing: while she made an informed and instrumentally rational choice (given her values), she chose to return to being someone who merely does Dianne's bidding. This requires her to once again return to the child-like state of being emotionally manipulated by Dianne and susceptible to her logical fallacies.

Think of Jesse, who is addicted to benzoylmethylecgonine. The pleasure that cocaine provides him is so intense that he does not want to give it up. Jesse is aware of the drug's side effects; he continues to use it just because of the way it makes him feel. If Jesse were to become fully informed and fully rational, he would still desire to use cocaine; like Molly, Jesse values his own happiness above all else. Full-information DSW deems Jesse best off continuing to be addicted to cocaine because that is what he most desires (even under idealized conditions). Intuitively, this seems wrong, however: although Jesse is fully informed, he seems to have acted against, not for, his own interest.

I contend that the same holds true for Molly. Intuitively, I believe that we think there is something wrong here, that even though Molly made an informed and instrumentally rational choice, it turned out to be the wrong one: she was better off choosing to leave the brainwashing.

We are inclined to think that although the happiness Molly receives from rejoining Dianne's ranks benefits her, that benefit is outweighed by the harm done to Molly by being brainwashed. Thus, while full-information DSW accords with intuition in the case of Sally, it does not in the case of Molly. What seems unacceptable about full-information DSW is that it is a view about what constitutes welfare. While it is clear that fully informed and fully rational individuals are in a better position to know what is best for them, it does *not* follow that what is best for them just is satisfying fully informed desires.

There is another worry that I have about full-information DSW: if full-information DSW is right, in order for a satisfied desire to be good for you, your fully informed and fully rational self must have the same desire. Given that a fully informed and fully rational Sally would not desire to continue her involvement with the cult, she would not desire her own happiness while a part of the cult. Yet Sally seems better off brainwashed and happy than brainwashed and miserable. Some of our desires seem better for us to fulfill even if a fully informed and fully rational version of ourselves would not want us to satisfy them.

Additionally, both pure DSW and full-information DSW are unable to accord with the intuition that happiness is *intrinsically* good for us. Full-information DSW maintains that happiness is good for us *only if* we would desire it while fully informed and fully rational. DSW also fails to accord with this intuition, since it maintains that something is good for you *only if* you desire it. A desire-satisfactionist about welfare could accept this criticism and just maintain that welfare consists of *both* satisfied desires and happiness. But if she makes this move, we no longer have a desire-satisfaction theory of welfare; rather, we have an objective list theory of welfare with two items on the list: desire-satisfaction and happiness. We shall discuss such theories below, in section 5.6. Briefly, while I think that such an objective list is an

improvement over DSW, since it better accords with intuition, I think that my own theory is superior. This brings us to the second theory of welfare we shall consider.

5.2 Life-satisfaction about welfare (LSW)

5.2.1 The essence of LSW

LSW is the position that the essence of welfare is the satisfaction of your life as a whole: you are well off to the extent that you are satisfied with your life as a whole, worse off to the extent that you are dissatisfied with it. Like HHW and DSW, LSW is a strong claim: it says that in all cases, prudential value is essentially determined only by life-satisfaction; the more satisfied S is with her life, the better off S is, period. You are satisfied with your life according to LSW in the same manner in which you are satisfied with your life according to LSH. That is, life-satisfaction, according to LSW, has the same two components that life satisfaction does for LSH: *cognitive* and *affective* satisfaction. To be cognitively satisfied with your life is to find that it on balance measures up to your standards and evaluations; to be affectively satisfied with your life is for you to find your life rewarding, for you to feel fulfilled by living it. Cognitive and affective satisfaction each are necessary for life satisfaction, together they are sufficient for it. Now that I have explicated the essence of LSW, let us see how well it fares with the case of brainwashing.

5.2.2 LSW and Sally

Not only does Sally have the most overall pleasure while brainwashed, not only does she have the most overall satisfied desires while brainwashed, she also is very satisfied with her life as a whole. In the first months after she leaves the cult, Sally is not at all satisfied with her life as a whole, since she is just now coming to grips with the fact that she was brainwashed. Several

years after she leaves the cult Sally is satisfied with her life, indeed, significantly more satisfied than she was only a month or two after leaving the cult yet not nearly as satisfied as she was while she was in the group.

If this is correct, the life-satisfaction Sally has while brainwashed, by itself, makes her all-things-considered better off. While I doubt that Sally's life-satisfaction by itself makes her better off, let us assume for the moment that it does. If her life-satisfaction benefits her, the benefit, I maintain, is *merely pro tanto*: once we consider what brainwashing is, the benefit life-satisfaction provides Sally with is outweighed by the harm that the brainwashing does to her. Even though she has more life-satisfaction brainwashed than months or even years after she is no longer brainwashed, Sally seems better off years after the brainwashing, all things considered. If this is correct, LSW is false.

A proponent of LSW could accept the above criticism and yet modify her theory. She could appeal to authenticity in the same manner the proponents of authentic HHW and full-information DSW appealed to authenticity, by maintaining that it is *authentic* life-satisfaction that constitutes your welfare.

5.2.3 Authentic LSW

The proponent of authentic LSW deems Sally's brainwashed life satisfaction *inauthentic*. At first glance, this seems plausible. Sally is being brainwashed, after all. Even though she believes that she is living a life fully of her own accord, she is being transformed into Dianne's slave: she is being deceived and manipulated into a lifestyle that if she had all the right information and was acting rationally, she would not choose to practice.

But what, precisely, does the proponent of authentic LSW mean by 'authenticity'? According to authentic LSW's leading proponent, Wayne Sumner, for S to adopt ϕ authentically

is for S to adopt her *own* self, free from social conditioning (Sumner 1996: 167). Authenticity, for Sumner, has two components: *full information* and *autonomy*. Sumner recognizes that full information is *insufficient* for authenticity: “the problem here is rooted not in the adequacy of people’s factual information but in the malleability of their personal values” (Sumner 1996: 162). This is a rather apt observation; full-information DSW, after all, ran into trouble because a fully informed and fully rational Molly still *values* her own happiness above all else. Autonomy is intended to address the malleability of a person’s values; but Sumner finds none of the three different accounts of autonomy that he considers completely successful. I want to criticize the account that Sumner seems to think is the most plausible, an account by John Christman.

According to Christman, S adopts φ authentically if and only if S adopted φ fully knowing the nature and effects of φ or would adopt φ if she fully knew its nature and effects (1991: 10-13). He is concerned with the process by which S adopts φ : so long as S either adopts φ under the right process or would have adopted φ under the right process, S adopts φ autonomously. So, authentic LSW is the position that you are well off to the extent that you would be satisfied with your life as whole if you were fully informed. Although Christman speaks of ‘processes’, what determines whether S would adopt φ is whether she was fully informed. And remember: mere full information is not sufficient to render Sally worse off brainwashed than not: if at any time during the brainwashing Sally were to become fully informed, she would desire to stay brainwashed. She would see all of the alternative lives she has available and choose the one that continues her brainwashing because that is what she currently values most. (It is full information in combination with full rationality that makes a fully informed and fully rational Sally aware that she is being brainwashed.) There are similar problems for the other accounts of autonomy Sumner discusses. He acknowledges this, leaving

us with only the thought that autonomy has been violated when “autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, rolescripting, and the like” are present (1996: 171). Sumner leaves it to others to find the correct theory of autonomy.

While he leaves us without a viable theory of autonomy, he does maintain that authenticity consists, in part, of full information. Think back to full-information DSW. Authenticity there consisted of both full information *and* full rationality. Yet full-information DSW only accorded with intuition in the case of Sally, not the case of Molly. Let us see how full-information LSW fares. It maintains that Sally is well off to the extent that a fully informed and fully rational Sally would be satisfied with her life as whole. A fully informed Sally would see all of the consequences of her possible actions; and a fully rational Sally would be able to judge free of cognitive errors and the errors of instrumental rationality which life she would be most satisfied with. Given that a life free of brainwashing would better meet the goals and expectations of a fully informed and fully rational Sally more than a life of brainwashing, full-information LSW deems Sally better off not brainwashed. Thus, full-information LSW accords with intuition in the case of Sally. But what happens when we consider Molly?

Well, Molly, unlike Sally, desires happiness more intensely than anything else. Thus, she would be most satisfied with a maximally happy life. So a fully informed and fully rational Molly would be most satisfied with the life that brought her the most happiness. And because the life that brings her the most happiness requires her to be brainwashed, she would continue the brainwashing. Like full-information DSW, then, full-information LSW fails to accord with intuition in the case of Molly.

I want to return to Sumner's idea of the "autonomy-subverting mechanisms of social conditioning", like "indoctrination, programming, brainwashing, [and] rolescripting" (1996: 171). Suppose that life-satisfaction that is caused by these autonomy-subverting mechanisms does not contribute to your welfare. In the case of brainwashed Sally, since all of her life-satisfaction is the result of these mechanisms, authentic LSW deems her better off years after the brainwashing, in accordance with intuition. Yet it also deems Sally better off just a couple months after her brainwashing: while she has no authentic life-satisfaction while being brainwashed, she has some just a couple months after the brainwashing. Thus, this authentic LSW fails to accord with intuition, too.

There is one last modification I shall try. Think back to the authentic hedonisms I tried above in section 4.8. Rather than maintain that authentic happiness alone contributed to welfare, they maintained that happiness contributed to your welfare to the extent that it was authentic: authenticity was not a component of welfare; rather, it was something like a multiplier. Suppose that we do the same thing for LSW, formulating it as follows. According to authentic LSW, welfare consists of life-satisfaction, where a particular episode of life-satisfaction makes you well off to the extent that it is authentic. Authenticity, here, consists of full information and the absence of autonomy-subverting mechanisms. Authentic LSW accords with intuition in the case of Sally: although she has no authentic life-satisfaction while brainwashed, Sally has much life-satisfaction, enough to outweigh the benefit that the authentic life-satisfaction she has just a couple months after being brainwashed brings her. This modification also accords with intuition in the case of Molly: Molly may desire happiness above all else, even if it is inauthentic; yet authentic LSW deems her better off with the authentic life-satisfaction that she has years after leaving the cult.

While this final formulation of authentic LSW accords with intuition in the cases of Sally and Molly, it seems incomplete. According to it, one is worse off with life-satisfaction that arises from autonomy-subverting mechanisms than one is with life-satisfaction that does not arise from such mechanisms. But what are autonomy-subverting mechanisms? We can point to Dianne's manipulations of Sally as an example of autonomy-subversion; but what, *specifically*, is it about Dianne's being brainwashed that subverts Sally's autonomy? Authentic LSW only stipulates that brainwashing subverts our autonomy. A theory that does not just stipulate brainwashing subverts our autonomy but *explains* why is, other things equal, more plausible than authentic LSW. My own theory of welfare—capacity fulfillment—not only accords with intuition on the cases of Sally and Molly, it also provides an explanation for why brainwashing subverts our autonomy. Sure, authentic LSW could just adopt the explanation that my theory provides; but note that in doing so, LSW is becoming more like my own theory. Moreover, LSW—authentic or not—is also susceptible to the following criticisms that my own theory is not susceptible to.

5.2.4 Additional criticisms of LSW

I want to question the idea that life satisfaction *by itself* makes Sally better off. Remember, to be satisfied with one's life as a whole is to be both affectively satisfied and cognitively satisfied with one's life. It is clear to me that pleasure seems intrinsically good for us; but I am less certain that affective satisfaction is. To be affectively satisfied with your life is to find it enriching, rewarding, or fulfilling. It is not clear to me that the *mere* feelings of enrichment, reward, or fulfillment are prudentially valuable. If these feelings are *pleasurable*, then, yes they seem prudentially valuable. But they seem prudentially valuable merely because of the pleasure they bring about; they do *not* seem prudentially valuable apart from the pleasure.

Imagine that you have just climbed the daunting 14,411 feet that is Mt. Rainer. At the peak of Rainer feelings of enrichment, reward, and fulfillment overwhelm you. You feel, in a word: great. While I of course admit that these feelings are good for you, they seem good for you because of the way that they make you feel. If this is right, it is not the feelings themselves but their *feeling good* that is good for you. Hence, as I see it, it is the *pleasure* you receive from these feelings that is good for you, not the feelings themselves.

Relatedly, both LSW and authentic LSW fail to accord with the intuition that happiness is intrinsically prudentially valuable. According to LSW, happiness is good for you only *in virtue of* it contributing to your overall satisfaction with your life; likewise, according to authentic LSW, happiness is good for you only in virtue of it contributing your overall life satisfaction. As we discussed above, happiness does not seem good for us merely instrumentally; rather, it also seems *intrinsically* good for us.

It is also not clear to me whether cognitive satisfaction—being satisfied that your life measures up reasonably well with your expectations—is *intrinsically* prudentially valuable, either. It seems to me that most of the time it feels good to find that our lives do measure up reasonably well to our expectations. When this happens I do not dispute that the pleasure you receive thereby makes you better off. But again, this would be because of the pleasure, not cognitive satisfaction alone. I could even agree that if you desired that your life match up with your expectations, it matching up with your expectations would be good for you; but this would mean that your cognitive satisfaction is good for you *in virtue of* you desiring it. I just doubt that it is the combined state of affairs of your reflecting and finding that your life on the whole worthwhile that by itself makes you any better off.

That said, it seems obvious that life-satisfaction is an *indicator* of welfare: generally speaking, a person who is not at all (either cognitively or affectively) satisfied with her life as a whole does not seem likely to be well off. Given the above discussion, however, the relationship between life satisfaction and welfare seems merely *contingent*: life satisfaction—authentic or not—just does *not* seem to be intrinsically prudentially valuable. If this is correct, LSW is false.

5.3 Self-fulfillment theory of welfare (SFW)

5.3.1 The essence of SFW

SFW is the position that the essence of welfare is self-fulfillment: you are well off to the extent that you fulfill your self, worse off to the extent that you do not fulfill your self.⁷¹

Obviously there are two parts to SFW that need to be explicated: the self and fulfillment. We shall start with the former first, then move on to the latter.

The notion of the self that is relevant for SFW does *not* concern the notion of reidentification of individuals over time (Haybron 2008a: 183). That is a question for those primarily interested in personal identity. Rather than be concerned with *which* individuals we are, the advocate of SFW is concerned with *who* we are; this is a question about our “thicker” selves (Haybron 2008a: 183-184). And our thicker self, Haybron contends, has both a rational side *and* an affective side. He admits that it is difficult to state what precisely the components of either side are. Nonetheless, Haybron maintains that the rational side consists of your projects, commitments, values, beliefs, and self-conception. And the affective side, given his emphasis on emotions where happiness is concerned, must consist of affective states and moods.

⁷¹ I take SFW from Haybron’s 2008 text. Strictly speaking, Haybron does not endorse SFW: he acknowledges that well-being may not just consist “in the fulfillment of the self’s two parts, but also in the fulfillment of our subpersonal, “nutritive” and “animal” natures: physical vitality and pleasure” (2008a: 194). In personal correspondence, Haybron confirms that well-being, in his view, does not merely consist of self-fulfillment.

Now that we have an idea of what the self is in self-fulfillment, let us discuss fulfillment.

Fulfillment consists of three *necessary* elements; together, all three are *sufficient*.

Enjoyment: φ fulfills S's nature only if she *enjoys* performing φ (Haybron 2008a: 175).

Authenticity: φ fulfills S's nature only if she *authentically* performs φ (Haybron 2008a: 186).

Internalism: φ fulfills S's nature only if by performing φ she fulfills the rational or affective sides of her individual self (Haybron 2008a: 157).

Thus construed, according to SFW, φ is valuable for S if and only if she authentically enjoys performing φ and if by performing φ she fulfills the rational or affective sides of her individual self. That is, if SFW is right, S is well off to the extent that she authentically enjoys fulfilling the nature of her individual thicker self. All three of these claims require further explication.

I began this dissertation with a discussion of happiness and pleasure, and I continued the discussion throughout the dissertation. Clearly, pleasure is related to well-being; it seems to be intrinsically prudentially valuable. The question is just to what extent it is prudentially valuable. SFW not only maintains that pleasure is good for you, it maintains that in order for φ to be intrinsically good for you φ needs to bring you pleasure. That is, if SFW is correct, the relationship between pleasure and welfare is one of *necessity*: pleasure is a necessary component of welfare. Now we need to discuss internalism.

Internalism, Haybron tells us, embodies two ideas.

First, what counts toward my well-being must depend on what I am like. My welfare must not be alien to me, a value that floats down from some Platonic realm and, remora-like, affixes itself to me with little regard to the particulars of my constitution. Second, what counts toward my well-being must not depend on what any other individual, or group or class of individuals—actual or hypothetical—is like. It must be possible to specify the ultimate or fundamental conditions for my well-being without making essential reference to other individuals, or to classes or groups of individuals. (2008a: 157)

Haybron's internalism constraint grounds welfare in *your* nature, not anything alien to you, but who you essentially are; no external authorities decide what is best for you. As a part of SFW, internalism implies that φ fulfills S's nature only if by performing φ she fulfills her projects,

commitments, values, beliefs, or self-conception or she fulfills her affective self. S seems to fulfill her affective self, according to SFW, to the extent that she is happy. Remember, happiness for Haybron consists of positive central emotional and mood states. Now that we have discussed both enjoyment and internalism, let us move on to authenticity.

Haybron maintains that authenticity has five “dimensions”: 1) being well enough informed, 2) autonomous values, 3) autonomous activities, 4) an autonomous temperament, and 5) richness (2008a: 186). Richness applies to happiness: the more complex the sources of your happiness are, the more “the authenticity of one’s happiness increases” (Ibid.). Unfortunately, he does *not* elucidate upon what autonomy *is*, what distinguishes, say, autonomous values from non-autonomous values. Indeed, he admits that his account of authenticity is “somewhat cursory” and is best left vague since you may simply insert whatever theory seems most plausible to you (2008a: 186).

I contend that the most plausible account of authenticity involves successfully using your critical capacities, which in turn involves not only on being well enough informed but also rational; being rational consists of avoiding errors of instrumental rationality and cognitive errors. Thus understood, you are rational, in part, to the extent that you are not susceptible to logical fallacies and emotional ploys. This closely resembles full-information DSW, but it is different in one crucial respect: it makes welfare, at least in part, dependent upon how well you reason, not on the desires of one who reasons well. So unlike full-information DSW, this account deems Molly better off not being brainwashed since she better exercises her critical capacities free of the brainwashing. But note that SFW does by becoming more like my own theory of welfare, capacity-fulfillment, which maintains that you successfully exercising your capacities, including your critical capacities, makes you better off.

5.3.2 SFW and Sally

Now that I have briefly explicated SFW, we need to discuss its verdict on the case of Sally. Sally is being brainwashed by Dianne, that is, she is being transformed into a mere deployable agent of Dianne's. Sally's self-identity and self-conception have changed: where she once identified herself as a loyal daughter and successful undergraduate with ambitions for going to medical school, she now identifies herself as nothing but an extension of Dianne. Not only has the manner in which Sally regards herself changed, her deepest values, commitments, and beliefs have all changed, as well. If she were to take a personality type test, her personality type would now be identical with Dianne's.⁷² Indeed, the brainwashing leaves Sally no longer able to cognitively or affectively deindividuate herself from Dianne.

Haybron wants to say that Sally is better off not being brainwashed. But remember: Sally no longer has the same self. Her deeper rational and affective selves have changed as a result of the brainwashing. When measuring her self-fulfillment we do *not* measure the self-fulfillment Sally had before or after the brainwashing; rather, we must measure the self-fulfillment in her current *brainwashed* self. If we do this, brainwashed Sally's life seems high in self-fulfillment, in fact, the highest it has ever been: never before has she to this extent lived up to her values, beliefs, commitments, emotional nature, and so on. Remember, intuition is clear, here: Sally while being brainwashed seems better off not brainwashed. SFW agrees that Sally before she is brainwashed is better off not becoming brainwashed and agrees that Sally years after the brainwashing is better off not returning to the cult to be brainwashed again; but SFW does not accord with the intuition that brainwashed Sally is better off not brainwashed.

⁷² See Yeakley (1985) for empirical research that members of destructive cults change their personalities in conformity to the identity of their leader.

5.3.3 SFW's responses to Sally

Haybron has a ready response to this objection. He of course emphasizes the *authenticity* constraint: φ fulfills S's nature only if she exercises φ authentically. Brainwashed Sally clearly does not exercise, well, anything authentically while brainwashed; she is brainwashed, after all!—or so says the proponent of SFW. Let us consider this claim.

If we interpret autonomy as successful use of one's critical capacities, then it is clear that SFW deems Sally better off not brainwashed. If Sally were not being emotionally manipulated and fooled by logical fallacies, she would not have the values, temperament, and so on that she possesses. So SFW deems Sally not at all well off brainwashed. But here is the catch: SFW deems brainwashed Sally *too* badly off. While Sally seems worse off brainwashed than years after the brainwashing, she seems better off brainwashed than just month or two after the brainwashing. Yet if SFW is correct, Sally would be worse off brainwashed than a month after the brainwashing. Thus, SFW fails to accord with intuition in the case of Sally.

There is at least one other response that a proponent of SFW could make. She could maintain that some of Sally's possible thicker selves are better for her to fulfill than others. While it may be better for brainwashed Sally to fulfill her self than not fulfill it, she would be even better off fulfilling a self that is not brainwashed. I think that this is right: other things equal, if Sally never leaves the brainwashing, Sally seems better off fulfilling the values and beliefs that her self-conception consists of than not fulfilling them; it just so happens that brainwashed Sally seems even better off having a different self-conception, say, the self-conception she has years after the brainwashing, and fulfilling the values and beliefs that that self-conception consists of. What SFW, then, requires is a mechanism that permits it to *rank-order* Sally's possible selves. The problem is: how does SFW rank-order the selves when it

views each individual self as merely the contents of one's beliefs, desires, and affections? To start, SFW would need to give up its internalism constraint. SFW seems to err in not going deep enough. It needs some sort of tool to delve *deeper* into Sally that can say of a certain thicker self that it is better or worse for Sally to fulfill it than another. The best way to do this, I think, is for SFW to ground welfare not in Sally's fulfillment of her self, but in her fulfillment of her capacities. If we do this, however, we wind up moving closer to my own theory of welfare.

SFW, in addition to failing to accord with intuition in the case of Sally, faces additional criticisms.

5.3.4 Additional criticisms of SFW

As I see it, the three elements SFW deems necessary for welfare—authenticity, enjoyment, and internalism—do *not* seem necessary for welfare. I shall start with authenticity.

SFW's authenticity claim seems false. If it were true, *inauthentic* happiness would be of no intrinsic prudential value; but *inauthentic* happiness does seem intrinsically prudentially valuable: Sally seems better off brainwashed and happy, all else equal, than brainwashed and unhappy. Contra SFW, then, authenticity does *not* seem necessary for welfare.⁷³

Nor does the enjoyment claim seem true: while pleasure seems prudentially good, in order for ϕ to be good for you, it does not appear that you need to enjoy ϕ . Think back to Daniel the ascetic from 2.1.3 above. He satisfies many of his intense desires, yet his life consists of hardly any pleasure. Sure, he seems unhappy; but at least he gets what he wants. Daniel seems better off, other things equal, unhappy with many of his intense desires satisfied than unhappy with few of his intense desires satisfied.

⁷³ Haybron seems to respond to something like this concern about SFW when he suggests that pleasure has prudential value apart from self-fulfillment (2008a: 94). So while SFW requires authenticity, self-fulfillment does not exhaust welfare, according to Haybron.

I also do *not* agree with SFW's internalism constraint: while Sally fulfills what Haybron would consider the rational side of her self extraordinarily well, she seems better off not fulfilling it. That is, if brainwashed Sally were to fulfill her projects, commitments, values, beliefs, self-conception, and so on, she would stay brainwashed as long as she could; but she seems better off leaving the cult, the sooner the better. It does not seem necessary, then, that for φ to make S better off φ needs to fulfill S's projects, commitments, values, and so on.

Not only do I contend that authenticity, enjoyment, and internalism do not seem to be necessary components of welfare, I also question whether SFW is justified in emphasizing happiness: why think that S's affective self is fulfilled by her being happy, not her being unhappy? It seems that you fulfill the affective dimension of your individual self just as well by being unhappy as you do by being happy: both happiness and unhappiness consist of central emotions and moods. I wonder what makes positive central emotions and moods fulfill your self but not negative ones. This brings us to our next theory of welfare to consider.

5.4 Perfectionism about welfare (PW)

5.4.1 The essence of PW

PW is the position that welfare consists of the perfection of the nature of your species: you are well off to the extent that you perfect your nature qua member of the species homo sapiens, worse off to the extent that you do not. Notice that this makes PW different than SFW, which grounds your welfare not in your nature qua member of homo sapiens, but in your nature qua individual self. PW has a rich pedigree, with proponents like Aristotle, Kant, Marx, and Nietzsche. Proponents of PW, generally speaking, define the term 'nature' in two different ways. First, following Aristotle, a thing's nature is that which it *distinctly*, uniquely, possesses. Second, following more recent perfectionists, like Thomas Hurka, a thing's nature is the set of its

essential properties. I explicate and criticize both views below before describing how PW needs to be modified in light of my criticism.

5.4.2 Aristotle's PW

Aristotle seems to think that the nature of a living thing is that which makes it unique or *distinctive*. The nature of a human, then, does not consist of being alive, since all living things are alive. Nor does a human's nature consist of activities like growing and digesting: plants and animals engage in these activities, too (*Nicomachean Ethics*: 1097b25-1098b5). Our nature may also not consist of perception, since, non-human animals perceive, too. The one *unique* activity we do, Aristotle famously states, is *rational* activity. That which something uniquely does is its *function*. Following the function argument, a good human is someone who performs her function excellently.⁷⁴ He then concludes by saying that the good of a human consists in excellent rational activity.⁷⁵ According to Aristotle, then, you are well off to the extent that you perfect your rational abilities, worse off to the extent that you fail to do so.

Now, the Greeks of course lacked the English 'welfare'. While they did use 'eudaimonia', it is not clear to me that it and 'welfare' are synonymous. 'Eudaimonia', unfortunately, is often translated as happiness. As discussed above, 'happiness' is ambiguous. On the one hand, it may mean just a particular psychological, and therefore, *descriptive*, state. On the other, it may mean the sort of life that is *good for* the one leading it. Given that we already have several clear terms for the latter sense of 'happiness', it is best to reserve 'happiness' for the former sense. Aristotle clearly did *not* mean the cheery, fell-good,

⁷⁴ As Hurka notes (1993: 10), Plato (*Republic*: 353a) also seems to think that a thing's good lies in its function.

⁷⁵ This interpretation of Aristotle is not without controversy; some disagree with it. See, for example, Kraut (1979). According to Kraut, Aristotle concludes that the good of a human consists in virtuous activity, be it intellectual or moral.

descriptive state of happiness; rather, he meant some sort of valuable state. Etymologically, ‘eudaimonia’ means having a good guardian spirit.⁷⁶ While it is clear that to achieve or possess eudaimonia is good or valuable, it is not clear to me that it is just good for you. That is, while it is clear that eudaimonia is a type of value, I am not sure that it only tracks prudential value. A life that is high in eudaimonia may be a life that is good for the person leading it; but it may not be just that: a life full of eudaimonia may be a life that is all-things-considered good, not just prudentially good. Construed in this way, a life full of eudaimonia would be a life high not only in prudential value but moral and aesthetic (perhaps other value), too.⁷⁷ Setting this concern aside, let us presume that ‘eudaimonia’ tracks *only* a particular type of value: *prudential* value.

Clearly, this version of PW deems Sally better off years after the brainwashing than while brainwashed: while brainwashed she fails to exercise her rational capacities well at all. In fact, Dianne is only able to brainwash Sally by dulling her reasoning. Sally exercises her rational capacities better both a month after the brainwashing and a several years after the brainwashing. Thus, while Aristotle’s PW accords with intuition in deeming brainwashed Sally worse off than she is several years after the brainwashing, it, conflicts with intuition in deeming Sally worse off brainwashed than a month after the brainwashing. In the month or two after leaving the cult Sally is distraught. She realizes that she has been brainwashed and does not know who, if anyone, to trust. She therefore feels rather miserable much of the time and even occasionally entertains her own suicide. Sally seems better off brainwashed.

⁷⁶ I get this definition from *The Oxford Companion to Philosophy*. Taylor, C. C. W. "eudaimonia." The Oxford Companion to Philosophy. : Oxford University Press, 2005. Oxford Reference. 2005. Date Accessed 11 Oct. 2013 <<http://www.oxfordreference.com.www2.lib.ku.edu:2048/view/10.1093/acref/9780199264797.001.0001/acref-9780199264797-e-812>>.

⁷⁷ See, for instance, Haybron (2008) who understands eudaimonia in this way.

Not only does Aristotle's PW fail to accord with intuition in the case of Sally, it also fails to give any prudential weight to happiness: if something does not involve exercising your reason, it makes you no better off whatsoever. Hence, this version of PW fails to accord with the intuition that happiness by itself has at least some benefit for the happy. Sally seems better off brainwashed and happy, all else equal, than brainwashed and unhappy. Aristotle's PW, however, deems her *equally* worse off brainwashed and happy as brainwashed and unhappy, miserable, even.

There are other objections to Aristotle's PW, too. Peter Glassen focuses on the function argument. If Glassen is right, Aristotle conflates the good of a thing with a thing's good. Aristotle describes what it is to be a good lyre player, that is, what the goodness of a lyre player consists in; but he wants to give us what the good of a lyre player consists in (Glassen: 320-321). So at best what Aristotle gives us, if Glassen is right, is what it is to be a good human, not what our good consists in.

Suppose that we revise the function argument in light of Glassen's worry. That is, suppose that we replace Aristotle's premise *a good human is someone who performs her function excellently* with *the good of a thing consists in a thing exercising its function excellently*. Now we seem to have a valid argument. The thing is: is it sound? That is, is this a true premise? I do *not* think so: why should we think that a thing's good must be something it *uniquely* does? Suppose we learn that Martians exist and reason precisely like we do; then reason would no longer be unique to us, and hence, reason could not be our good. Why should the facts that Martians exist and reason like us impact what it is that makes us well off?⁷⁸ The idea that our good is determined by facts about other species seems implausible.

⁷⁸ Hurka (1993: 11) and Robert Nozick (*Philosophical Explanations*: 515-517) make the same point.

Further, following Bernard Williams, even if we are the only creatures who reason the way we do, there seem to be many abilities we uniquely possess that do *not* seem relevant for our welfare. Using Williams's examples, it seems that we are the only beings who make fire, have sex "without regard to season", "despoil the environment", and kill for fun (Williams 1972: 59). None of these activities by themselves seem to make us better off in the slightest.

5.4.3 Hurka's PW

I think it is safe to say that due to the above criticisms Aristotle's PW does not seem plausible. Fortunately for the perfectionist there are other versions of PW. Instead of defining nature in terms of distinctness, as Aristotle does, one could define it in terms of *essence*. This is the move that Thomas Hurka makes. Taking a page from Saul Kripke and Hilary Putnam, Hurka thinks of ϕ 's essence as its essential properties, which understood modally, are those properties it has in every possible world (where it exists). Now, Kripke and Putnam, for the most part, only discuss the essential properties of physiochemical kinds at the micro level. The essence of gold, for example, they argue, is the atomic number 79: in every possible world where gold exists, while it may not be yellow, shiny, and so on, it will have the same atomic structure denoted by atomic number 79. Hurka goes a step further in applying essential properties to humans. An *essential* property for a member of the homo sapien species, then, is a property that a human would have in *every possible world*.

But there is a problem with this proposal that Hurka recognizes immediately: it, like Aristotle's notion of essence, is guilty of the wrong properties objection. There are some properties that we possess essentially that do not seem relevant for our welfare.⁷⁹ For example,

⁷⁹ Strictly speaking, Hurka is not a proponent of PW: while he is a perfectionist, he does *not* maintain that perfectionism tracks welfare (Hurka 1993: 17). So while he maintains that you live a good life to the extent that you perfect your human nature, he does *not* maintain that such a life is *good for* you; such a life is just good, period.

in all possible worlds we are identical with ourselves. Self-identity is, then, an essential property. But it would be very odd to think that being self-identical somehow constitutes your welfare. So he further maintains that we must limit essence to properties that are “*conditioned on their being living things*” (Hurka 1993: 16). The property of being identical to yourself is not a property conditioned on you being a living thing; rather, it is a property conditioned on you being an object. Hurka aptly refers to his theory as an “essence-and-life” view (Hurka 1993: 17).

What properties are essential to humans qua living things? First off, humans, Hurka says, have an essential *physical* nature, constituted by circular, digestive, muscular, nervous, and respiratory systems. If something is not able to “breathe, move, process nutrients, and exercise central control, it is not a human” (Hurka 1993: 37). Well, then, what is it to perfect one’s physical nature? Hurka’s answer is simple: achievement of good bodily health in that the aforementioned bodily systems “function in an efficient, unrestricted way” (Hurka 1993: 38). There seems to be some intuitive support for this. Think of a soldier who goes off to war healthy, with her bodily systems operating efficiently. Suppose that when she returns, however, she is without his right leg. Lacking her leg, she does not seem complete. She seems less well off, not just because she may get less pleasure and less of what she wants; the loss of her leg itself seems like it harms her.

Humans do not only have a certain physical nature; according to Hurka, they also have *practical* and *rational* natures, too. Humans are practical in that we not only “have local aims”, we also “can envisage patterns of action that can stretch through time or include other agents and can perform particular acts as a means to them” (Hurka 1993: 39). Humans are rational in that

Welfare, in his mind, is subjective, consisting of desires. I take his theory of perfectionism and use it as a kind of theory about welfare.

we “can form and act upon sophisticated beliefs and intentions, ones whose contents stretch across persons and times and that are arranged in complex hierarchies” (Hurka 1993: 39).

You perfect your *rational* nature, Hurka tells us, to the extent that you are *disposed* to have justified true beliefs that are highly general (Hurka 1993: 99-101).⁸⁰ Thus, it is not just the *number* of justified true beliefs you are disposed to have, it is also their *generality* that is relevant for your welfare. On Hurka’s view, “It is better to know a fundamental law of the universe than the number of redheads in Beiseker, Alberta, or the workings of a friend’s personality than the exact length of his forearm” (Hurka 1993: 100). A scientific law is general in the sense that “it includes all the objects at all times in history” (1993: 115). But beliefs, Hurka insists, may be general in another sense: they may also be general if they explain other true beliefs (Ibid.). I think the basic idea is that if belief b is more general than belief c, b is necessary in some sense to explain c; b is explanatorily prior to c.

If our rational capacities really are essential to us, then, Hurka maintains, they “must be realized to some degree at every time” of our lives (Hurka 1993: 101). And if they “must be realized to some degree at every time”, they must be *dispositional*, since dispositional beliefs “persist through sleep and unconsciousness” (Hurka 1993: 101). Hurka also contends that his dispositional account of the rational good is more intuitive than a merely occurrent account: “Imagine that one person knows one truth, which he contemplates at every time in his life, whereas another knows many truths that she contemplates in succession, one after the other” (Hurka 1993: 101). A perfectionist theory of welfare that emphasizes only occurrent beliefs would deem the first person’s life as just as valuable as the second’s, which is counterintuitive.

⁸⁰ Hurka (1993: 105-106) relies on an internalist account of knowledge as justified true belief. If this makes you uncomfortable, I see no reason why you cannot substitute an externalist account of knowledge as, say, true beliefs based on reliable causal mechanisms.

You perfect your *practical* nature, Hurka tells us, to the extent that you are *disposed to intend* to achieve your goals that you are *justified* in thinking will be achieved (Hurka 1993: 101-103). Intending, for Hurka, is different than mere desiring: where desires are “idle wishes”, intentions are wishes someone has “set herself to pursue” (Hurka 1993: 102).⁸¹ The extent to which a particular intended achievement makes you better off is determined by the quality of the achievement itself. On his view, “Someone who through elaborate planning achieves a major reform of his society does more of intrinsic worth than someone who merely ties a shoelace” (Hurka 1993: 100).

Now that we have discussed the specifics of Hurka’s essentialist PW, let us see how it fares with the case of Sally. The relevant differences between Hurka’s essentialist PW and Aristotle’s distinctive PW is that in addition to including rational perfection, Hurka’s account also includes physical and practical perfection. Sally is less physically healthy just a month or two after the brainwashing than she is either brainwashed or several years after she leaves the brainwashing. In the first couple of months after she leaves the cult, Sally feels depressed: she does not sleep as well, eat so well, or exercise much. During the brainwashing she treats her body much better; and several years later she is back to treating it well again. Sally also has lots of practical perfection while brainwashed. While some of her intentions are unjustified—like, her intention to save the world—in the sense that they are not realizable, many of her intentions are justified. She intends to be the second-in-command of the cult, and she intends to lead many to the group. Thus, she exercises her practical nature much better while brainwashed than the first couple of months after the intervention. Hurka’s PW, then, deems her better off

⁸¹ Keller (2009) makes a similar distinction between goals and mere desires: goals aim at achievement whereas mere desires do not (670-672). If you have a goal to achieve ϕ , Keller maintains that you are thereby “commit[ted] to bringing about the targeted state of affairs” (672). Whereas if you merely desire ϕ , you are *not* thereby committed to bringing about ϕ .

brainwashed than a couple months after the intervention. And even though years after the brainwashing Sally does not perfect her practical nature quite as well as she did while brainwashed, she perfects her rational nature to a much greater extent. Hurka's PW, then, deems Sally better off years after the brainwashing than being brainwashed. Thus, Hurka's PW, unlike Aristotle's, accords with intuition in the case of Sally.

While it accords with intuition in the case of Sally and is thus an improvement over Aristotle's PW, Hurka's PW faces numerous challenges. First off, the capacities Hurka maintains are essential—namely, the particular physical, practical, and rational capacities—do *not* seem essential. Think, for example, of a fetus. It certainly does not possess circulatory, respiratory and other physical systems, let alone practical and rational capacities. Hence, it is not a member of the human species, on Hurka's view. Same thing with certain of the mentally or cognitively impaired: because they lack the rational capacities Hurka deems essential to human beings, they are not human beings. This does not seem plausible, however. Think of when someone who is cognitively impaired dies, we think that a human being has died, not a non-human.⁸²

Like Aristotle's PW, Hurka's PW fails to give any intrinsic prudential worth to pleasure. Richard Arneson has us think of cheap thrills, activities that “provide pleasure and excitement without any significant effort or sacrifice on the part of the agent and also without the exercise or development of any of the agent's significant talents” (Arneson: 11-12). These pleasures seem to enhance our lives while they offer no “redeeming social value beyond their pleasantness” (Arneson: 12).

If the above section is correct, neither Aristotle's or Hurka's PW seem plausible. Not only does Aristotle's PW fail to accord with intuition in the case of Sally, his function argument

⁸² Kitcher makes the same point (Kitcher 1999: 69).

seems unsound and his account of nature deems certain of our capacities relevant for our welfare when they do not by themselves seem relevant. As for Hurka's PW, while it accords with intuition in the case of Sally, it also deems certain of our capacities essential to us that do not seem essential. Moreover, neither PW grants any intrinsic prudential worth to pleasure.

Let us try to construct a version of PW that is not susceptible to the criticisms that Aristotle's and Hurka's PW are susceptible to. They both seem correct to emphasize the rational part of our selves. It enables them to deem that Sally is better off years after the brainwashing than brainwashed. As we saw, Aristotle's account exclusive reliance on rationality emphasizes it too much: while we are inclined to think that Sally is better off brainwashed than just a couple months after her brainwashing because of how miserable she is, Aristotle's PW deems her better off miserable and not brainwashed. Hurka's account, however, is not entirely free of this worry because it, too, fails to accommodate the idea that happiness is intrinsically prudentially valuable.

In response, then, I think that a proponent of PW should not only contend that rationality is intimately connected with our welfare but our happiness, too. Any perfectionist account of welfare grounds our welfare in our natures. The two views given above—Aristotle's uniqueness and Hurka's essentialism—fail. What we need is either an account of nature that does not fail, or a theory of welfare that makes no appeal to nature whatsoever. Thus far, I have suggested that happiness and satisfied desires are intrinsically good for us. If we add successfully using one's critical capacities to the mix, we can accord with intuition in the case of Sally. Why, then, not just suppose that welfare consists of these goods? If we go this route, we no longer have a perfectionist theory of welfare, but an objective list, which we shall now discuss.

5.5 The objective list (OL)

5.5.1 The essence of OL

OL is the position that the essence of welfare is the collection of certain goods: you are well off to the extent that your life is full of the certain goods, worse off to the extent your life does not include these goods. These goods, whatever they are, are good for you *independent* of your attitudes or feelings toward them. This is what makes them *objectively* good. A typical OL includes items such as: achievement, friendship, knowledge, and pleasure. The list of the aforementioned goods is said to be objective because it permits something to be good for you independent of your feelings or attitudes toward it: your autonomy, achievements, and knowledge make you better off even if you are not positively disposed to them.

What I would like to do now is explore what OL says about Sally.

5.5.2 OL and Sally

To determine whether OL accords with intuition in the case of Sally, we need to know which goods are on the objective list. Let us suppose that this particular objective list contains the following aforementioned goods: achievement, friendship, knowledge, and pleasure. During the brainwashing Sally achieves many of her wants and experiences much pleasure. She also has considerably more knowledge of what is in the Bible and how to influence people; it is just that she is not very knowledgeable about how destructive her church is. She does make many new friends who are also members of her group, but she loses several friends who are frightened away by her new lifestyle; this leaves her without a net gain or loss in the number of her friends. And while she feels especially close with her new friends in the group, she is no longer as close with several of her friends outside of the group; this leaves her without a net gain or loss in terms of how close she feels with her friends.

A month or two after the brainwashing Sally is more knowledgeable about the church and brainwashing; but she often feels lousy and achieves little of what she desires. At this time, then, Sally has more knowledge; but she now has a dearth of two of the goods she had in spades while brainwashed: achievement and pleasure. She loses more in terms of achievement and pleasure than she gains knowledge. After leaving the cult, Sally has left behind many of her friends that remain in the group; yet she is closer with her friends outside of the group. This seems to leave Sally no better or worse off in terms of her friendships before or after she leaves the cult. Thus, according to this particular objective list Sally is better off brainwashed than she is a month or two after the brainwashing. Years after the brainwashing, though, Sally experiences much pleasure, is much more well informed about the group and the nature of brainwashing, and achieves much of what she wants. Still, she does *not* have as much pleasure nor get as much of what she wants as she did while brainwashed. The gain in knowledge she has does not outweigh the decrease in her pleasure and achievement. Thus, this objective list deems her better off brainwashed, even than years after the brainwashing, in conflict with intuition.

The proponent of OL may not think that this is a problem. After all, she may just insist that while this particular objective list conflicts with intuition in the case of Sally, not all objective lists do; and the right one, the proponent could insist, accords with intuition. We just have to find it. Well, at least two items on the previous list seem intrinsically beneficial to Sally: achievement and pleasure. So let us leave them on the list. This time around, I shall drop friendship. And in addition to knowledge, I shall add critical thinking, which is being devoid of cognitive errors and the errors of instrumental rationality.

During her brainwashing Sally receives much achievement and pleasure. She has a moderate amount of knowledge about the Bible and how to influence people. However, she has

a dearth of critical thinking: she is victim to many logical fallacies and emotional ploys. A month or two after leaving the brainwashing, Sally does not have much achievement or pleasure. She does have more knowledge and critical thinking; it is just that her gain in these two items falls short of the lost achievement and pleasure. Several years after the brainwashing, however, while Sally experiences less achievement and pleasure, she has significantly more knowledge and critical thinking; so much so that her gain in these two items outweighs her lost achievement and pleasure. This particular objective list, then, deems Sally best off years after the brainwashing, in accordance with intuition.

This objective list also deems Molly better off not being brainwashed, too. Sure, she most intensely desires her own happiness and thus wants to continue the brainwashing. Nevertheless, this objective list deems brainwashing bad for Molly since it violates her critical capacities. Years after the brainwashing, although Molly has less pleasure, she successfully exercises her critical capacities to a much greater extent. Thus, this objective list deems her better off years after the brainwashing, in accordance with intuition.

5.5.3 Support for OL

Unlike all of the above theories of welfare except HHW, this particular objective list accords with the intuition that happiness is intrinsically good for the happy. According to it, happiness is a constituent of well-being. Thus, every episode of happiness makes you better off. But unlike HHW, this objective list does not deem that happiness is the only constituent of well-being; rather, it also maintains that achievement and critical thinking constitute well-being, too. Further, unlike all of the above theories of welfare except DSW, this particular objective list accords with the intuition that satisfied desires are intrinsically good for us. According to this objective list, getting what you want is another constituent of well-being. So while it implies that

every satisfied desire makes you better off, this objective list, unlike DSW, does not maintain that satisfied desires are the only things that make you better off. This objective list also takes a lesson from the above theories of welfare by implying that exercising your critical capacities free of error is intrinsically good for you.

While this particular objective list accords with intuition in the cases of Sally and Molly and has considerable strengths, objective list theories of welfare face the following two criticisms. I do *not* find the first to be convincing but the second I do.

5.5.4 Criticisms against OL

Objection 1: OL violates the internalism constraint

Objective lists fail to respect the internalism constraint, which basically maintains that for φ to be intrinsically good for you, you must be motivated to want φ , at least under ideal conditions.⁸³ Desire-satisfaction theories of welfare accord with this constraint: if you desire something, you are by definition motivated to want that thing. If this constraint is correct, something cannot be good for you without your say-so. This may seem rather plausible: it is *your* welfare we are talking about, after all. And who else but you is in the position to know whether φ makes *you* better off.

Remember, OL maintains that φ may be good for you independent of any motivations that you may have. Think about what this implies. Suppose that your life contains the goods listed on the above objective list. The thing is: you are repulsed by these parts of your life. According to OL, though, your life is going well for you, in fact, very well for you. It is no surprise, then, that without the internalism constraint critics accuse OL of being paternalistic, of

⁸³ See Rosati (1996: 298) for a full explication and defense of the internalism constraint.

restricting your liberty against your will for your supposed good. Internalism gives you the authority to determine what is good for you. I do not find this criticism convincing, however.

I think that the internalism constraint is implausible. Think back to brainwashed Sally. She is motivated to do only one thing: promote the interests of Dianne. And if Sally pursues her motivations, she seems to be pursuing her own peril: she seems better off acting against her motivations by leaving the brainwashing than she seems acting in accordance with her motivations. Sally's motivations are the product of, amongst other things, misinformation and a failure to exercise her critical capacities. The particular objective list we discussed that accords with intuition in the case of Sally does not make Sally's well-being subject to the whims of anything like typical motivations are subject to. While OL may alienate you from your current desires, wants, projects, and so on, it may—depending on the particular objective list—prevent your good from consisting in the satisfaction of motivations that you would abandon based on fuller information and rationality. In a sense, a given objective list may permit us to better get at Sally's *deeper* self, the self not susceptible to misinformation and cognitive errors.

While I do not think that the internalism constraint offers a good criticism of OL, I think that the following objection does.

Objection 2: OL needs an explanation

This theory of welfare is different from the others that we have discussed. Consider hedonism. If asked to create a list of items that are prudentially valuable, the hedonist may come up with the following: achievement, enjoyable experiences, and knowledge. But if asked what makes the items on her list good for us, the hedonist would say that it is their being pleasurable. This is because the hedonist insists that pleasure is the *only* thing that is intrinsically good for us. Thus understood hedonism is not merely an *enumeration* of prudentially valuable goods; rather,

it is an explanation for why those goods are prudentially valuable.⁸⁴ That is, hedonism is *not* merely a list of the sources or components of well-being; rather, it is an account of the nature of well-being. This is as it should be. A theory of welfare is not only an enumeration; it is also an *explanation* for why those items are prudentially valuable. The OL, however, is just that: a list of items that purport to be good for you. We need not only a list, but an explanation. In failing to provide an explanation, OL theories fail to even be theories of welfare, properly speaking; or so it is argued.⁸⁵

Proponents of OL have a response in waiting for this. They maintain that, despite the name, OL is *not* merely a list of purported prudential goods; it is also an explanation for why those goods are good for you.⁸⁶ Unlike hedonism, which maintains that only *one* thing is intrinsically good for you, OL maintains that *multiple* things are intrinsically good for you. Suppose a proponent of OL maintains that autonomy, knowledge, and pleasure are the only things that are intrinsically prudentially valuable. If this is right, ϕ makes you better off to the extent it brings you autonomy, knowledge, or pleasure. The theory just adds autonomy and knowledge in addition to pleasure. That is, according to OL, there is not only one thing that is good for us; rather, several things are intrinsically good for us.

I think that this response misses the thrust of the criticism. A simple enumeration of purported prudential goods does not seem to be as theoretically unified as a theory that purports that there is only one thing that is good for us. The proper response a proponent of OL should give, in my eyes, is as follows. While one may prefer a theory with just one item for its

⁸⁴ Crisp makes the same distinction (2006: 622-623).

⁸⁵ Sumner makes a similar argument (1996), as does Valerius (2008).

⁸⁶ Arneson, for one OL theorist, argues in this manner (1999: 9-10).

theoretical unity, a proponent of OL could maintain that her theory better accommodates our commonsense beliefs about prudential value. And we would rather have a theory that accommodates our intuitions but is not theoretically unified than one that fails to accommodate our intuitions but is theoretically unified.

Notice that this final response leaves OL susceptible to a theory that accommodates intuition well as it does that also is theoretically unified. This is where I think that the criticism holds true. Take any two theories of welfare, A and B; if they accord with intuition as well as each other and do not have any internal inconsistencies but A is more theoretically unified than B, A is the more plausible theory. All of the above theories claim to be not only a list of prudential goods but an explanation for why those items are good for you; but each fails to accord with intuition as well as this particular objective list does or it is susceptible to criticisms that this objective list is not. Thus, while this particular objective list may not be as theoretically unified, it seems more plausible. If there were, however, a theory that accommodates intuition as well as this objective list that is theoretically unified, well, that theory would be more plausible than this objective list. I think that I have found such a theory. I shall explicate it in the next and final chapter.

Chapter 6: What Welfare Is

In this chapter I introduce my own theory of welfare, show its strengths, and then defend it from criticism.

6.1 Capacity-fulfillment theory of welfare (CFW)

6.1.1 The essence of CFW

I shall first formulate CFW as the position that the essence of welfare is the fulfillment of your capacities: you are well off to the extent that you fulfill your capacities, worse off to the extent that you fail to do so. The idea of capacity-fulfillment, of course, requires some explication. I shall first discuss what a capacity is, next what fulfillment is.

Capacities are best understood as *powers* or *abilities*. You have the power to bench-press 50 pounds; you have the ability to add two and two together. We never just possess a capacity, simpliciter; rather, we possess a capacity *for* φ . That is, capacities are relational. They are also *numerous*: we have capacities for many things. I could be writing my dissertation right now, or I could be playing a video game; I have the capacity for either. We do not, however, possess capacities for that which it is impossible for us to do. While I may wish that I could run at the speed of light, that is something I cannot do and, hence, do not possess the capacity for. But you *can* possess a capacity for φ without occurrently exercising φ . For instance, when you fall asleep, you are not, generally speaking, exercising your rational capacities. Yet you still possess your capacity for rationality: you are just not occurrently being rational. Now that I have said a few things about what a capacity is, let us discuss fulfillment.

I understand fulfillment as follows: φ fulfills a capacity c of S's only if by performing φ she exercises c . My idea of fulfillment, unlike Haybron's, then, entails neither authenticity nor enjoyment: you may fulfill a capacity of yours even if you get *no* enjoyment from doing so and

even if you exercise the capacity *inauthentically*. You may also fulfill a capacity without being aware that you are doing so. Say you are an avid fan of *The New York Times* crossword puzzle. Every morning the first thing that you do after getting out of bed is turn the stove on in order to heat up your tea kettle. As you wait for the water to boil, you open up the *Times* on your iPad and start the daily crossword puzzle. While completing the puzzle, you exercise numerous capacities, including your capacity to complete the *Times* puzzle. As you are completing the puzzle it is not as if you are thinking, ‘Boy, I really enjoy exercising my capacity to complete the *Times* crossword puzzle’. No, you need *not* be aware that you are exercising a capacity in order to exercise that capacity; just by completing the *Times* crossword puzzle you are exercising your capacity for doing so.

But clearly *not* all capacities are created equal: if we suppose that fulfilling a capacity makes you better off, fulfilling some capacities seem to make you better off than fulfilling others. For example, not only do you possess the capacity to complete the *Times* crossword, you also possess the capacity to count blades of grass. Other things equal, it seems better for you to complete the crossword than count blades of grass.

I can accommodate this concern in several ways. First, I could take a page from the desire-satisfactionist’s playbook and maintain that S is well off, at least in part, to the extent that she fulfills the capacities that a fully informed and fully rational S would want the actual S to fulfill. There is an obvious problem for this formulation, however. Remember, full-information DSW failed to accord with intuition in the case of Molly. A fully informed and fully rational Molly would want brainwashed Molly to be maximally happy. So long as a fully idealized Molly values happiness above all else, she would commit no cognitive errors nor errors of instrumental rationality in wanting actual Molly to be as happy as she can be. Likewise, a fully

informed and fully rational Molly would want actual Molly to fulfill the capacities necessary for her maximal happiness; if being maximally happy requires her to be brainwashed, then so be it.

So, it does *not* appear that the idea of full information helps me differentiate between capacities that are better for Sally to fulfill than others. I could take a page from the perfectionist's playbook, maintaining that fulfilling capacities that Sally *naturally* possesses is better for her than fulfilling capacities that she does not naturally possess. If I go this route, how do I define 'nature'? I argued above that Aristotle is wrong to define 'nature' in terms of distinctiveness and that Hurka is wrong to define it in terms of essence. If my arguments are successful, these two notions are of no help. We could, however, just drop the species requirement and instead focus on the *individual*, as Haybron does. Thus construed, your essential capacities are not those essential to you qua human being but those essential to you qua *individual self*. So your essential capacities would be those that you possess in virtue of being *you*. This, too, is problematic, however: there seems to be some possible world where you exist without the capacity for affection. So the idea of essential capacities qua human being or qua individual does not seem to aid us. There is something that does, however, help us distinguish between those capacities that seem relevant for our welfare from those that do not: *basic capacities*.

6.1.2 Basic capacities

In order to explicate what a basic capacity is, I think it is useful to refer back to DSW. First note that desire-satisfactionists distinguish between *intrinsic* and *instrumental* desires: only the former, they maintain, contribute to welfare. Suppose that you desire both to drink an expensive bottle of wine with your friends and that you desire to impress your friends; but you only desire to drink the wine in order to impress your friends. A desire-satisfactionist about

welfare will say that only the desire to impress your friends if fulfilled increases your welfare: your desire to drink the wine is merely an instrumental desire; it has no prudential value for its own sake.

I make a similar move with CFW. In exercising some capacities you are exercising others. And it is only the fulfillment of basic capacities that constitutes your welfare.⁸⁷ A basic capacity is a capacity not explained by any other capacity; a non-basic capacity is a capacity that is explained by some other (basic) capacity. Your capacity to complete the *Times* crossword puzzle is *not* a basic capacity, since it may be explained by other (basic) capacities. So, what are the basic capacities? As Shidan Lotfi sees it, there are four: affection, agency, cognition, and conation.⁸⁸

Affection consists, foremost, of feeling pleasure (positive affect) and pain (negative affect). We also possess the capacity to feel emotions, though it seems to be the case that “[e]very emotion is associated with pleasantness or unpleasantness” (Lotfi: 68). Agency consists of forming intentions and carrying those intentions out. While other animals may intend to do things, only humans, Lotfi tells us, can form “complex goals and pursue long-term plans” (Lotfi: 70). Cognitive capacities are those that you use to *perceive* and *represent* the world. Hearing, tasting, feeling, and so on are sense capacities that enable us to perceive and represent the world. We also have the ability to reason. The direction of fit is of world to mind in the sense that you

⁸⁷ Lotfi (2011) also relies on the idea of basic capacities. But Lotfi does not merely rely on basic capacities; he relies on basic *essential* capacities. According to Lotfi, the capacities that are relevant for our welfare are not only basic, they are essential, too. For the reasons explicated above, however, essential capacities seem problematic. Rather than rely on essential basic capacities, I just rely on basic capacities. And as Lotfi notes, he borrows his distinction between basic and non-basic capacities from Kit Fine and Michael Gorman, who distinguish between basic and non-basic essential properties. Lotfi takes Fine’s and Gorman’s distinction between properties and applies it to capacities. According to Fine, “An essential property of an object is a constitutive part of the essence of that object if it is not had in virtue of being a consequence of some more basic essential properties of the object; and otherwise it is a consequential part of the essence.” (1995: 57) According to Gorman, “F is essential to x just in case F is (i) a characteristic of x and (ii) not explained by any other characteristic of x. By contrast, F is accidental to x just in case F is (i) a characteristic of x and (ii) explained by some other characteristic of x” (2005: 284).

⁸⁸ Again, strictly speaking, Lotfi maintains that there are four basic *essential* capacities.

fit your mind or beliefs to the world. We exercise our conative capacities when we *desire*. The direction of fit is the opposite of cognition, namely: mind to world in the sense that you fit the world to your mind or desires. When you desire something, you want to make the world a certain way. There are multiple levels of desire. You may desire that you obtain a tenure-track position next year. You may also desire to have the desire to obtain a tenure-track position next year. The former desire is a first-order desire whereas the latter is a second-order desire.

I think that Lotfi errs by distinguishing between agency and conation. As I see it, intending is a form of wanting, that is, desiring. Perhaps intending is a more intense form of desiring; but this is a difference in degree, *not* in kind. If your desire for φ involves long-term plans and complex goals but your desire for ψ does not, your desire for φ will likely be more intense than your desire for ψ . So while satisfying your desire for φ will, other things being equal, make you better off than satisfying your desire for ψ , this is because it is a more intense desire, and not because of intending having intrinsic prudential value. The carrying out of one's intentions I would call 'volition', a part of conation.

If this is right, there are just three basic *mental* capacities: affection, cognition, and conation. The idea that we possess three basic mental capacities has a rich pedigree. Plato divided the human psyche (soul) into three parts: epithumia (appetite), logos (reason), and thumos (passion).⁸⁹ Kant described the human mind as consisting of the capacities to desire, feel, and know.⁹⁰ The fact that the likes of Plato and Kant seemed to have thought that we

⁸⁹ See *Republic*, especially Books 4, 8, and 9, and *Phaedrus*.

⁹⁰ See *Critique of the Power of Judgment* and *Anthropology from a Pragmatic Point of View*.

possess three basic capacities lends the idea some initial plausibility.⁹¹ But I wonder whether these three are all the basic capacities that we possess.

Let us think back to Sally. Throughout much of her life she has treated her body reasonably well, working out three to four times a week. One of her favorite exercises at the gym is the benchpress. I want to determine the basic capacities that she is exercising while benching. She wants to keep her body in good shape both because she wants to be a good steward of her body and because she just likes the way feeling in shape feels. While she is benching she feels pain; but she also takes pleasure in the strengthening of her muscles. Thus far, she is exercising two basic capacities: affection and conation. How do we explain her physical exertion, though? She not only feels some pain and pleasure; nor does she just fulfill some of her desires; no, she is also moving her body. Hurka seems right to include a physical dimension to his theory of welfare. I would like to import it. So, as I see it, there are four basic capacities: *affection*, *cognition*, *conation*, and *locomotion*.

Now, we do not only need a distinction between a basic capacity and a non-basic one. Otherwise, what makes fulfilling your capacity for pleasure better for you than fulfilling your capacity for pain? Surely, it is better for you to be happy, other things equal, than unhappy; yet in both states you would be fulfilling a basic capacity: affection.

I think that *successfully* fulfilling your basic capacities is good for you. You successfully fulfill your basic capacities when they are functioning properly.⁹² You successfully fulfill your

⁹¹ Lotfi makes the same point (62-63).

⁹² See Lotfi and Simon Keller for other theories of welfare that rely on the notion of success. Lotfi maintains that you are well off to the extent that you successfully fulfill your basic capacities (76-85). Keller maintains that you are well off to extent that you are successful, where to be successful is “to have attitudes that do well according to the standards they constitutively set for themselves” (674). There are four such attitudes according to Keller: beliefs, goals, evaluative attitudes, and pleasure (666-674).

affective capacities whenever you experience pleasant feelings or pleasant emotions. Unlike SFW, then, CFW explains why happiness, not unhappiness, constitutes affective fulfillment. Think of normal functioning human beings. Do they prefer pleasure to pain or pain to pleasure? Clearly, they prefer pleasure to pain. There is no need to adopt psychological hedonism, which implies something like the idea that we are only motivated to intentionally pursue acts that bring us more pleasure than pain. Rather, all that we need to say is that, generally speaking, psychologically healthy individuals pursue what they think will bring more pleasure than pain. If someone seeks pain over pleasure, we typically think that she is suffering from some sort of affective, psychological, or affective-psychological disorder.

You successfully fulfill your *cognitive* capacities to the extent that you represent the world *accurately*, free from error. This involves accurate sense perception, that is, accurate feeling, hearing, seeing, smelling, and tasting. Successfully fulfilling your cognitive capacities also entails reasoning free from error, emotional ploys, fallacies, and errors of instrumental rationality. It also consists of having knowledge. Suppose you believe that Socrates was the teacher of Plato, but you do not know: you heard a friend of yours mention this once; but this friend of yours has a reputation of frequently being wrong. While this true belief alone represents the world accurately and thus, according to CFW, increases your welfare, *knowing* that Socrates was the teacher of Plato is a more successful use of your cognitive capacities and thus, according to CFW, increases your welfare more.

While I agree with Hurka's PW that knowledge is intrinsically good for us, I do *not*, however, agree that one piece of knowledge is more intrinsically prudentially valuable than another. No matter how general any two bits of knowledge are in comparison to each other, they seem *equally* intrinsically prudentially valuable. Of course I admit that some bits of knowledge

increase your welfare more than others. Bits of knowledge that you take pleasure in or enable you to satisfy a desire are more prudentially valuable than those that do not. Note, though, that the increased prudential value here is *instrumental*, not intrinsic: there is nothing about some knowledge itself that makes it more prudentially valuable; it is the pleasure or satisfied desire that you have in addition to the knowledge that makes one bit of knowledge more prudentially valuable than the other. That said, of course bits of knowledge that enables us to gain more knowledge are, other things equal, more instrumentally valuable than those that do not, too.

Refer back to Hurka, who says “It is better to know a fundamental law of the universe than the number of redheads in Beiseker, Alberta, or the workings of a friend’s personality than the exact length of his forearm”. I do *not* think that knowing a fundamental law of the universe is by itself better for S than knowing the number of redheads in Beiseker. Sure, there may be some beliefs that it is better for S to know than others. But this is not in virtue of some quality—like generality—that some beliefs possess. Rather, it is a result of some beliefs enabling S to fulfill her capacities better than others. So while knowing a more complex belief may contribute to your welfare more than a less complex one, we need not rely on Hurka’s generality to get it.

My theory of welfare also differs from Hurka’s PW in that it is *neutral* in regards to what constitutes knowledge. Hurka’s PW, remember, entails an internalistic understanding of knowledge where S knows *p* if and only if *p* is true, S believes *p*, and S is justified in believing *p*. Justification, here, is an internal manner: S is justified in believing *p* only if her evidence for believing *p* makes *p*’s truth probable. While Hurka’s analysis of knowledge may be correct, there are notorious problems with knowledge conceived in the aforementioned manner.⁹³ Rather than get bogged down in adjudicating between internalism and externalism about knowledge, I remain neutral. CFW, then, makes a similar move in regards to knowledge as HH makes in

⁹³ See Gettier (1963) and Goldman (1976), amongst others.

regards to pleasure. As HH maintains that pleasure constitutes happiness without committing itself to a particular understanding of pleasure, CFW maintains that welfare is in part constituted by knowledge without committing itself to a particular understanding of knowledge. Just take whatever theory of knowledge you find most plausible; having more of that, CFW maintains, is good for you, less of it, bad for you.

I do, however, agree with Hurka that capacity-fulfillment is best thought of as *dispositional*. Think of two individuals who are both unconscious, say, sleeping. The only difference between the two is that the first is disposed to have more justified true beliefs, act free from rational errors, experience more pleasure, and have more desires satisfied than the second. The first seems better off than the second even though they are not occurrently acting free from rational errors, and so on.

Since your conative capacities consist of both desiring and volition, you successfully fulfill your *conative* capacities by getting what you want. I think satisfaction, here, should be understood *objectively*: successful desire-satisfaction occurs when you desire ϕ and ϕ . When we measure the extent to which a satisfied desire contributes to your welfare we must take into account the desire's *intensity* or strength and its *duration*. The more intensely and the longer you desire ϕ , the more satisfying your desire contributes to your welfare.

What is it to successfully exercise your *locomotive* capacities? As Hurka indicates, it is to be physically healthy, having your bodily systems “function in an efficient, unrestricted way” (Hurka 1993: 38). There is some intuitive support for this. Think of two individuals S and T. S does not take care of himself. His diet consists of lots of red meat, foods high in trans fat, and sweets. And he never exercises; he spends the time he could be exercising in front of his TV, munching on his potato chips. T is the opposite: his diet not only consists of healthy foods like

fish, nuts, and greens, he exercises about an hour a day. Now, both S and T roughly get the same amount of pleasure and both satisfy roughly the same number of desires. And as strange as it may sound, they are about as rational as each other. Even though they share about the same amount of pleasure, desire-satisfaction, and rationality, T seems better off than S.

The notion of *successful* capacity-fulfillment, here, accords well with our intuitions on well-being. Let us think about what it means to be affectively well off. Clearly, we tend to think that it means something like feeling good as opposed to feeling bad. What is it to be cognitively well off? It seems to mean something like being free from cognitive errors and having, in general, a preponderance of knowledge. To be conatively well off seems to consist of getting what you want. And to be physically well off seems to just consist of being healthy.

If CFW is right, you are just well off to the extent that you are well off in all four basic capacities. Happiness, satisfied desires, successful use of your reasoning or sense-perception, and good bodily health all intrinsically make you better off. Yet ϕ need not be pleasurable, need not be a satisfied desire, need not be a successful use of your cognitive capacities, nor successful use of your locomotive capacities in order for it to by itself make you better off; ϕ only needs to be one of those things for it to by itself make you better off. In order to be living a life that is good for you, you just need to be well off. In order to be well off, according to CFW, you need to have a positive overall amount of capacity-fulfillment. CFW is, then, *aggregative*. You are well off to the extent that you have a positive overall amount of capacity-fulfillment.

Now that I have finished explicating CFW, it is time to see whether it accords with intuition in the case of Sally.

6.1.3 CFW and brainwashed Sally

Sally exercises her affective capacities successfully while being brainwashed.

Remember, never has she felt so spiritually uplifted, so alive with power, so at ease with people, so confident of the truth of her convictions, and just so fulfilled. Not only does she feel great: Sally also takes pleasure in feeling so great. Indeed, she feels so pleasant and takes so much pleasure in feeling pleasant that Sally has a higher surplus of pleasure than pain, in fact, the highest surplus she has ever had over a similar duration. Sally's feeling happy is a successful use of her affective capacities. Because she has so much happiness, CFW deems her *affectively* well off brainwashed.

She also exercises her conative capacities successfully while brainwashed. Many of her intense desires go fulfilled, few go unfulfilled. Indeed, so many more desires of hers go fulfilled than unfulfilled that she has more fulfilled desires over the duration of her life while she is brainwashed than any other similarly long duration. Sally's getting what she wants is a successful use of her conative capacities. Because she has so many satisfied desires, then, CFW deems her *conatively* well off brainwashed, too.

Although Sally exercises her affective and conative capacities successfully while being brainwashed, she fails to exercise her cognitive capacities well. Indeed, she fails to represent the world well at all. Sure, her sense perception works fine; but Sally commits a plethora of cognitive errors. Remember, Sally has many true beliefs while she is being brainwashed; this is why truth-adjusted attitudinal hedonism fails to deem her worse off than years after the brainwashing. The problem with Sally is not that she does not have true beliefs but that she does not successfully employ her critical capacities while forming her beliefs. Dianne is only able to

get Sally to change her beliefs by manipulating Sally to reason poorly.⁹⁴ Her dependence on Dianne reduces her to a child-like state, leaving her rational capacities go unfulfilled. Thus, while CFW deems Sally affectively and conatively well off brainwashed, it does *not* deem her cognitively well off.

A month or two after the brainwashing Sally is now exercising her rational capacities in ways she was not while being brainwashed. Although she exercises her rational capacities more successfully, she often feels lousy and goes without getting much of what she intensely desires. Thus, while CFW deems her cognitively better off than when she was brainwashed, CFW deems her significantly affectively and conatively worse off. Indeed, while the successful exercise of her cognitive capacities makes her better off, according to CFW, her failure to exercise her affective and conative capacities leaves her with less overall capacity-fulfillment than she had while brainwashed. According to CFW, then, Sally is worse off a month or two after the intervention than she was while brainwashed. Thus far, then, CFW accords with intuition.

Several years after the brainwashing Sally is in a much better state, according to CFW. Not only does she exercise her rational capacities to the best she ever has, she is also predominately happy and satisfies many of her intense desires. It is just that Sally is not as happy nor satisfies as many of her desires as she did while brainwashed. Years after her brainwashing Sally has a large increase in capacity-fulfillment of her cognition and slight losses in her affection and conation as compared to when she was being brainwashed. This leaves her with more overall capacity-fulfillment than she had than while being brainwashed. Of the three periods, then, CFW deems Sally best off years after the brainwashing, in accordance with intuition.

⁹⁴ Keller's welfare as success is susceptible to this criticism: according to it, Sally is better off brainwashed than years after she leaves it since she has more true beliefs, fulfilled goals, true evaluative attitudes, and pleasure.

I contend that the problem with Sally is that she achieves her happiness at the *cost* of her rationality. While Sally is exceedingly affectively well off, she is not at all cognitively well off. She would be better off if she fulfilled more of her rational capacities. And if fulfilling more of her rational capacities requires her to leave some of her affective and conative capacities unfulfilled, so be it, so long as the result is a net increase in Sally's overall capacity-fulfillment.

Think of Mary, a supremely intelligent mathematician. She reasons thoroughly well and is very well informed. It is just that she is also thoroughly *unhappy*. Her problem is the opposite of Sally's: Mary achieves her mathematical prowess at the cost of her emotional health. While Mary is exceedingly cognitively well off, she is not at all affectively well off. She is better off fulfilling more of her affective capacities. Since fulfilling her rational capacities to the extent that she does prevents her from fulfilling her affective capacities, she is better off fulfilling fewer of her cognitive capacities, thus enabling her to fulfill her affective ones. This leaves Mary with more overall capacity-fulfillment.

While it is good that CFW accords with intuition in the case of Sally, several other theories discussed above did, as well: full-information DSW, sufficiently modified authentic LSW, Hurka's PW, and some objective lists. Full-information DSW, however, failed to accord with intuition in the case of Molly. What is CFW's verdict? Well, a fully informed and fully rational Molly, according to CFW, would be better off leaving the cult. Even though Molly desires her happiness above all else, CFW maintains that she is better off with the most overall successful capacity-fulfillment, not necessarily the most successful affective-fulfillment. And because she better successfully fulfills all four of her basic capacities outside the cult than within, CFW deems Molly better off years after the brainwashing, in accordance with intuition.

Now that I have explained why CFW accords with the cases of Sally and Molly, I shall explicate the strengths that CFW has over its rivals.

6.1.4 Strengths of CFW

1. CFW does not rely on an authenticity condition.

Unlike LSW and SFW, CFW does *not* require the messy notion of authenticity; so CFW does not get into trouble trying to provide an account of it. Sumner and Haybron both acknowledge that happiness that arises from brainwashing is less prudentially valuable than happiness that does not arise from brainwashing. Yet they have difficulty constructing an authenticity condition that deems brainwashed happiness prudentially defective. CFW has no such difficulty, since it implies that Sally's being manipulated by Dianne's emotional ploys and fallacies by itself harms Sally. In fact Sumner's LSW and Haybron's SFW are only able to deem brainwashing bad for those undergoing it *if* they maintain that reasoning free from error constitutes authenticity. CFW maintains not only that reasoning free from error is good for you, it also has an explanation for why it is good for you: reasoning from error is a part of successfully exercising your cognitive capacities; and since cognition is a basic capacity, successfully exercising it makes you better off.

2. CFW grants that enjoying ϕ is good for you but also that you need not enjoy ϕ for it to be good for you.

Unlike DSW, LSW, SFW, or PW, CFW also grants the apparent intrinsic prudential value of happiness. Whenever you experience happiness, you also experience pleasure (since happiness essentially is pleasure); and whenever you experience pleasure you exercise your affective capacities or your conative capacities (or both). While CFW permits any episode of pleasure to make you better off (in proportion to the intensity and duration of the pleasure), CFW does *not* go too far, like HHW and SFW, in making enjoyment a necessary condition for well-

being: according to CFW, the fulfillment of your basic capacities is good for you even if you receive no pleasure from fulfilling them. Someone who is successful and knowledgeable but who does not enjoy his success or knowledge is, all else equal, better off than someone without that success and knowledge; but CFW maintains that he is better off yet taking pleasure from her success and knowledge. Thus, CFW takes the best from hedonism (about welfare) since it implies that pleasure is good for you, that any episode of pleasure you experience makes you pro tanto better off. Yet CFW leaves out the bad from hedonism since it implies that things besides pleasure are good for you.

3. Like LSW, CFW has affective and cognitive dimensions to our welfare.

While happiness seems to merely be an affective matter, welfare seems to have both *affective* and *cognitive* dimensions. Welfare not only seems to consist of feeling fulfilled but also being fulfilled. LSW accords well with this. CFW is similar to LSW in that it includes both a cognitive and an affective dimension to welfare. Where authentic LSW maintains that one needs to be both cognitively and affectively satisfied to be well off, CFW maintains that one is well off to the extent one exercises her affective, cognitive, conative, and locomotive capacities successfully. Where LSW measures cognitive satisfaction in terms of how well you think your life meets your expectations, the latter measures cognitive well-being, at least in part, in terms of how informed and rational you are.

4. CFW has the benefits of full-information DSW without the cost.

CFW is not susceptible to the criticisms that full-information DSW is susceptible to. It takes the best—full-information DSW’s emphasis on information and rationality—and leaves out the troublesome part—the reliance on desire-satisfaction as a necessary condition for welfare. A fully informed and fully rational Molly values happiness above all else and therefore wants

brainwashed Molly to continue her brainwashing. This does not seem to be in her best interest, however: Molly seems better off years after the brainwashing, where she fulfills not only her affective and conative capacities, but her cognitive ones, too. Generally speaking, the more fully informed and rational someone becomes, the more likely she will know what is good for her. That said, being in such a state does *not* make someone infallible: one may still desire something that is not in her best interest. This accords well with CFW since it implies that becoming more fully informed and rational increases one's welfare.

5. CFW grants that satisfied desires by themselves benefit us.

CFW also accords with the intuition that satisfied desires by themselves are good for you. Take Seneca. He seems better off after delivering a rousing oration to the Senate even though he feels no emotional states. As I originally described it, Seneca enjoyed the fact that his desire to give a rousing oration was satisfied. Suppose this time that we take out the pleasure; that is, suppose that Seneca gets no pleasure from satisfying his desire. He still seems better off after giving the oration since an intense desire of his has been satisfied: sure, he would be better off still if he also received pleasure from doing so; but the satisfied desire alone seems to make him better off. And CFW accords with this. It just does not go too far, like DSW, in deeming desire-satisfaction a necessary component of welfare.

6. CFW does not rely on the questionable idea of human essence.

CFW, unlike PW, does not rely on essential or distinct properties; so it does not run into trouble distinguishing humans from non-humans.

7. CFW accords with intuition as well as OL; plus, it is theoretically unified.

My theory is also an improvement over the objective list containing pleasure, satisfied desires, knowledge, and reasoning well, since CFW accords with the world as well as this list of

goods and it has greater theoretical unity. As we discussed, OL faces criticisms for supposedly being an enumerative, not an explanatory, theory of welfare. CFW, unlike OL, does not just provide a list of prudential goods; rather, it provides a unifying explanation for why the goods it deems intrinsically prudentially valuable are intrinsically prudentially valuable. If CFW is correct, accurate sense perception, knowledge, rationality, pleasure, satisfied desires, and physical health are all intrinsically good for us. This is because each is the result of successfully fulfilling our basic capacities. Since CFW accords with the world as well as the aforementioned objective list, its greater theoretical unity makes it a more plausible theory of welfare.

Now that I have explicated CFW's strengths, I will attempt to defend it from criticism.

6.1.5 Defending CFW from criticism

1. Non-prudential desires criticism

DSW seems susceptible to the non-prudential desires criticism. Simply put, *some* satisfied desires do not seem to make us better off, even pro tanto. Refer back to the stranger on the subway that you spot in a cast. You desire of that person that he be free of that cast, that he be healed. Your desire is satisfied a week later when the person has a successful surgery, is healed, and hence no longer has need of the cast. It is clear that *he* seems better off without the cast. But how are *you* made any better off by his cast now removed? You do not even know about it.

While CFW does imply that any satisfied desire you have makes you pro tanto better off, the extent to which a particular satisfied desire makes you better off depends on the intensity and duration of the desire. So when you get things that you barely want, your welfare barely increases. Your desire for the person on the subway to be healed is not a very intense one, nor is it a long one; after all, you forget about it shortly after you get off the subway. So while

according to CFW the fact that the person is healed does make you better off, the extent to which it makes you better off is mitigated by how weak and short lasting the desire is. If your desire is only brief and fleeting, the satisfaction of it only has a brief and fleeting impact on your well-being.

Think of a related case involving Mary the mathematician discussed above in section 6.1.3. She never ceases trying to solve mathematical problems. Suppose that she does so because she believes that the axioms of logic and math are necessarily true. If it turns out the axioms of logic and math are merely contingently true, her life's work will be based on a lie. She intensely desires that the axioms be necessarily true; suppose that they are but that she does not know it. Like you on the subway, she desires something and it obtains without her realizing it. The difference is that an intense desire of hers has been fulfilled. In this case, I doubt that we are pretheoretically inclined to deem her no better off as a result of the desire being satisfied.

2. Trivial knowledge

CFW may be criticized for implying that all information you know makes you better off. Suppose that there are 50,000 words in Nietzsche's *The Birth of Tragedy*. Suppose further that you know this. By itself it seems that knowing how many words are in a book is rather *trivial* knowledge, knowledge that by itself has no relevance for your well-being. Yet if CFW is correct, you are better off in virtue of this piece of knowledge. At best, it seems that this knowledge is *instrumentally* good for you. Say that your knowledge of Nietzsche enables you to impress a cute girl at a bar who you later marry. Lucky for you, marrying her is really good for you. Knowing how many words there are in the *Birth*, then, was instrumentally good for you. But by itself, it seems trivial.

While CFW does imply that any bit of knowledge you possess makes you better off, certain bits of knowledge are better for us to possess than others: knowledge that enables you to successfully fulfill other basic capacities makes you better off than knowledge that does not. So knowledge that brings you pleasure or enables you to satisfy your desires is better for you, *instrumentally*, than knowledge that does not. That said, any true belief that you possess that is justified or based on a reliable causal mechanism (depending on your theory of knowledge), according to CFW, is just as intrinsically good for you as any other. I realize that this is counterintuitive; but I hope that the numerous strengths I described above outweigh this criticism.

Now, there may be a theory of welfare that accords with our commonsense beliefs as well as CFW but is not susceptible to the above trivial knowledge criticism. I am not aware of such a theory, however. One could argue that I have already discussed such a theory: the objective list with happiness, satisfied desires, and not being susceptible to emotional ploys and logical fallacies. But this is mistaken. While CFW is susceptible to the trivial knowledge criticism and this objective list is not, there is a type of case that CFW accords with intuition on that this objective list fails to accord with intuition on.

Think back to Ted and Bruce described in section 4.5.1.1. While both are equally blissfully happy businessmen, Ted's wife is cheating on him and he does not know it, whereas Bruce's wife is loyal; Ted's happiness is based on a lie, whereas Bruce's is not. Suppose that Bruce's belief that his wife is faithful is both justified and causally related to a reliable mechanism. And suppose that Ted's false belief that his wife is not the result of an emotional ploy or logical fallacy. Where Bruce, then, knows that his wife is loyal, Ted does not. Although they have equal amounts of happiness and success, and although they are equally

immune to erroneous reasoning, other things equal, Bruce seems better off than Ted. CFW deems Bruce better off than Ted because he has more knowledge than Ted, whereas the objective list that merely contains happiness, success, and immunity to emotional ploys and fallacies deems them equally well off.

The objective list that contains happiness, success, immunity to emotional ploys and fallacies, and knowledge does accord with intuition in this case. In fact, this objective list accords with our commonsense beliefs as well as CFW. The thing is: CFW is more plausible because it has an explanation for why the items it deems intrinsically good for us are good for us and this objective list lacks such an explanation

3. Harmful knowledge

CFW implies that knowledge is good for us, but clearly some knowledge is bad for us. Think about David, a 60 year-old man who has thus far lived a pretty happy and satisfying life. Unfortunately, he is having trouble with his vision lately. So he goes to see his eye doctor. The optometrist can find nothing wrong; so he recommends that David see his doctor. David's physician runs a battery of tests, including x-raying David's brain. He finds a mass. After getting it tested, they learn that it is a stage four malignant tumor. David becomes severely depressed once he learns this: he stops talking to others, loses his energy and vitality, and so on. A few weeks later he dies in the same state. Clearly, it seems that the knowledge made David worse off: it would have been better for him had he just not known about the cancer; he was going to die anyway. He might as well enjoy his last days in ignorance. CFW, however, implies that this knowledge is good for David. Hence, CFW seems wrong.

I have a ready response to this criticism. The proponent of CFW would say that David's knowledge of his tumor makes him *instrumentally* worse off: while the knowledge by itself

benefits him, according to CFW, that benefit is outweighed by the harm the knowledge does him in virtue of preventing him from successfully fulfilling affective and conative capacities that he would have otherwise fulfilled. A similar point can be made in the case of Sally: according to CFW, although brainwashed Sally's pleasure was intrinsically good for her, it was instrumentally bad for her since it prevented her from fulfilling her rational capacities successfully.

4. Accurate sense perception?

CFW maintains that successful use of your cognitive capacities constitutes a part of your welfare. And it also maintains that having accurate sense perception constitutes a part of successful cognition. This means that accurate sense perception is *intrinsically* good for you, which may seem rather implausible. Surely, one could argue, having accurate sense perception is at best *instrumentally* good for you. Take the capacities to hear and see. While these seem like important capacities to those who possess them, they seem valuable for us insofar as they enable us to, say, receive pleasure or satisfy our desires; by themselves these capacities do *not* seem valuable for us.

I am not so sure that intuition is clear on this one, however. Take Luke, who a few months ago went off to war. Unfortunately for him, he returns from the war with severe damage to his eyes, causing him to be blind. Luke's injury does seem instrumentally bad for him in the sense that (at least at first) he will no longer be able to do many of the things he once did. I think that we are pretheoretically inclined to say that Luke's injury *itself* harms him. Suppose that several years after returning home a new sort of surgery is performed on Luke that returns his sight. While being able to see again brings Luke much pleasure and desire-satisfaction, which seem beneficial to him in their own right, regaining the ability to see itself seems to benefit him, too.

Conclusion

I introduced this dissertation by discussing Sally, who seems happy, the happiest she has even been. The thing is: she is a member of a destructive cult being brainwashed to only pursue its ends. While Sally seems better off brainwashed than she does a month or two after she leaves the cult, Sally seems better off several years later, when she is thinking critically for herself again.

In chapter one I discussed some preliminary matters about happiness. I am looking for what it is that all and only people who are happy have in common. Happiness is different from welfare in that the concept of happiness is not value-laden like the concept of welfare. I would like to explicate the essence of what it is to be occurrently happy at a moment. I also discussed five commonsense examples of happiness: brainwashed Sally, Mary the mother, Glenda the graduate student, Seneca, and Andrew the swine.

In chapter two I explicated and then criticized three theories of happiness: DSH, LSH and EH. According to DSH, you are happy to the extent that your desires are satisfied. ‘Satisfied’ can be understood objective or subjectively. Understood objectively DSH seems implausible. Happiness is a psychological or mental state; if *objective* DSH is correct, however, you are made happier without any change to your mental states. And while *subjective* DSH accords with intuition in four of the five cases, it does not accord with intuition in the case of Glenda. Her intense desires remain just as unsatisfied after she takes the medication as before taking it; yet her mood dramatically changes: she goes from feeling lousy and depressed to feeling good and optimistic. With the mood-altering medication she is able to sleep at night and is much less irritable to others. Hence, we are pretheoretically inclined to judge that she is happier after taking the medication. DSH fails to accord with intuition here. While the happy may have more

of their desires satisfied than not, satisfied desires do not seem to constitute happiness: it seems possible for you to be happy even though your most intense desires go unsatisfied—as in the case of Glenda. And it seems possible for you to be unhappy even though your most intense desires go satisfied, as in the case of Daniel. Furthermore, subjective DSH has the strange effect that whenever you desire to be unhappy, you are the opposite of what you believe you are. Desire-satisfaction, understood objectively or subjectively, then, does not appear to be the essence of happiness.

According to LSH, you are happy to the extent that you are affectively and cognitively satisfied with your life. ‘Satisfied’ may be understood in two different ways here, too: actually or hypothetically. *Actual* satisfaction is clearly too strong: one can be happy without occurrently reflecting about one’s life. While *hypothetical* LSH accords with intuition in two of the cases, it fails to accord with intuition in three of the others, the case of Andrew in particular. While Andrew may enjoy disgusting things, he clearly seems happy. Yet if he were to reflect about the extent to which his life has lived up to his expectations, he would be dissatisfied. Hypothetical LSH thus deems him unhappy. It seems possible for one to be happy without being happy about her life. Thus, life-satisfaction, understood actually or hypothetically, does not appear to be the essence of happiness.

According to EH, you are happy to the extent that you have positive central affective states and positive mood propensities. While EH accords with intuition in four of the five cases, it fails to accord with intuition in the case of Seneca. He enjoys the state of affairs that includes his successful oration. Even though he fails to have any emotions or moods, Seneca clearly seems happy. Hence, positive emotions and mood propensities, contra EH, do not seem necessary for happiness. Mood propensities, in particular, seem unnecessary and insufficient for

happiness since actual moods, not mere propensities for moods, affect our happiness. Happiness also need not be central; it can be shallow and fleeting. Moreover, pleasure is what distinguishes positive emotional states from negative ones, making pleasure the more fundamental constituent of happiness.

In chapter three I explicated and then defended my own theory of happiness: HH. According to it, you are happy to the extent that you experience pleasure. HH differs from other hedonistic theories of happiness by remaining neutral about the nature of pleasure. Unlike the previous three theories of happiness, HH accords with intuition in all five cases. It also withstands the criticisms of Haybron. I contend that it is the most plausible theory of happiness.

In chapter four I explored the relationship between happiness and welfare. It seems obvious that happiness is good for you, that it is at least *pro tanto* beneficial to the happy. After explicating the nature of brainwashing, I used the case of Sally to show that happiness does not seem to be the essence of welfare: Sally seems better off not brainwashed even though she is happy. One could modify hedonism about happiness and welfare such that welfare is determined, at least in part, by *authentic* happiness. I provided two accounts of authentic HH: truth-adjusted attitudinal hedonism about happiness and desert-adjusted attitudinal hedonism about happiness. But they both fail to accord with intuition in the case of Sally, making them superfluous. Happiness—whether authentic or not, then, does not seem to constitute welfare.

In chapter five I first explicated and criticized five theories of welfare: DSW, LSW, SFW, PW, and OL. According to DSW, you are well off to the extent that your desires are satisfied. Satisfied desires do seem to be good for us: they seem at least *pro tanto* beneficial. Some desires, however, seem better for us to fulfill than others (independent of their intensity and duration). Even though Sally satisfies more intense desires brainwashed than not, she seems

better off, contra DSW, satisfying the desires she has years after the brainwashing. If we transform DSW into full-information DSW, the position that Sally is well off to the extent that she satisfies the desires that a fully idealized Sally would want her to want, it deems Sally better off years after her brainwashing, in accordance with intuition. Yet, this modification fails to accord with intuition in the case of Molly. I also discussed some additional problems with DSW and full-information DSW, the most important is that they only permit happiness to benefit you only if you desire it; whereas the benefit happiness brings us does not appear dependent on you desiring it.

According to LSW, you are well off to the extent that you are satisfied with your life to date. It does not accord with intuition in the case of Sally: because Sally has more life-satisfaction while being brainwashed than she does years after the brainwashing, LSW deems her better off brainwashed. Now, LSW can be modified to authentic LSW to accord with intuition in the case of Sally; but it can only do so by identifying authenticity as something like reasoning well, free from error. Hence, while LSW can be modified to accord with intuition in the cases of Sally and Molly, it only does so by becoming more like my own theory of welfare. Further, neither component of life-satisfaction appears to be intrinsically beneficial to us: any benefit that affective satisfaction brings you seems to stem from its pleasure, not affective satisfaction as such; and any benefit that cognitive satisfaction seems to come from getting what you want, not cognitive satisfaction as such. Moreover, LSW fails to accord with the intuition that happiness is intrinsically good for the happy.

According to SFW, you are well off to the extent you authentically enjoy fulfilling the rational and affective sides of your thicker self. SFW does not accord with intuition in the case of Sally because it deems her better off just a month after she leaves her brainwashing when she

is downright miserable than she is while brainwashed. In order to accord with intuition here, SFW seems to require a mechanism that can rank-order Sally's different thicker selves, deeming it better for her to fulfill the self she has years after the brainwashing than the self she has while brainwashed. The best mechanism to rank-order Sally's selves seem to be something like my idea of basic capacities. If we make this move, however, we no longer have a self-fulfillment theory of welfare but a capacity-fulfillment theory. Moreover, contra SFW, it appears that φ may make you better off without you exercising φ authentically or without you enjoying φ . Further, SFW's internalism requirement seems implausible: in order for φ to make you better off, it does not seem as though φ needs to fulfill S's projects, commitments, values, beliefs, and self-conception or affective self.

According to PW, you are well off to the extent that you perfect the nature of your species. 'Nature' can be understood either as what is distinctive of your species or as what is essential to it. Either way, however, essence is an implausible determinant of welfare: it either leaves us with a too restrictive definition of human being or it is susceptible to the wrong properties objection. Now, PW could be modified such that no longer points to a thing's nature. Instead, it could just list the goods that are intrinsically prudentially valuable. This is nothing more than an objective list, however.

According to OL, you are well off to the extent that your life contains the goods on the objective list. While the objective list that includes happiness, satisfied desires, knowledge, and reasoning well accords with intuition in the cases of Sally and Molly, it would be a more plausible theory of welfare if it were more theoretically unified. If a theory of welfare can be found that accords with the world as well as the aforementioned objective list while being more theoretically unified, that theory is more plausible. I contend that CFW is that theory.

This brings us to chapter six, where I introduce CFW, the theory that maintains you are well off to the extent that you successfully exercise your basic capacities. Not only does it accord with intuition in the cases of Sally and Molly, it is not susceptible to the criticisms that the above theories of welfare are susceptible to. That said, CFW does have two counterintuitive implications. It implies that any satisfied desire of your makes you better off and that any bit of knowledge you have makes you better off. I thus acknowledge that there are some tradeoffs with CFW; I just think that in the end I provide sufficient evidence for deeming CFW the most plausible of the welfare theories here discussed.

Bibliography

- Barker, C.H. and Edgar Schein. *Coercive Persuasion*. New York: Norton, 1961.
- Barrett, Lisa Feldman and James Russell. "The Structure of Current Affect: Controversies and Emerging Consensus". *Current Directions in Psychological Science* 8.1 (1999): 10-14.
- Benditt, Theodore. "Happiness". *Philosophical Studies* 25.1 (1974): 1-20.
- Bradley, Ben. "A Paradox for Some Theories of Welfare". *Philosophical Studies* 133.45 (2007): 45-53.
- Bramble, Ben. "The Distinctive Feeling Theory of Pleasure". *Philosophical Studies* 162.2 (2011): 201-217.
- Brandt, Richard. "Happiness". *The Encyclopedia of Philosophy*. Ed. P. Edwards. New York: McMillian Publishing Co, Inc. and The Free Press, 1967.
- Carson, Thomas. "Happiness, Contentment, and the Good Life". *Pacific Philosophical Quarterly* 62 (1981): 378-392.
- Christman, John. "Autonomy and Personal History". *Canadian Journal of Philosophy* 21.1 (1991): 1-24.
- Davis, Wayne. "A Theory of Happiness". *American Philosophical Quarterly* 18.2 (1981): 111-120.
- "Pleasure and Happiness". *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 39.3 (1981): 305-317.
- Den Uyl, Douglas and Tibur Machan. "Recent Work on the Concept of Happiness". *American Philosophical Quarterly* 20.2 (1983): 115-134.
- Dorsey, Dale. "Three Arguments for Perfectionism". *Nous* 44.1 (2010): 59-79.
- "The Hedonist's Dilemma". *Journal of Moral Philosophy* 8 (2011): 173-196.
- "Can Instrumental Value be Intrinsic?". *Pacific Philosophical Quarterly* 93 (2012a): 137-157.
- "Intrinsic Value and the Supervenience Principle". *Philosophical Studies*. 157 (2012b): 267-285.
- "Subjectivism without Desire". *Philosophical Review* 121.3 (2012c): 407-442.
- "Desire-Satisfaction and Welfare as Temporal". *Ethical Theory and Moral Practice* 16 (2013): 151-171.
- "Emotion and Consciousness". *Encyclopedia of Consciousness*. Oxford: Elsevier Science & Technology, 2009. *Credo Reference*.
- Eisenberger, Naomi. "Broken Hearts and Broken Bones: A Neural Perspective on the Similarities Between Social and Physical Pain". *Current Directions in Psychological Science* 21.42 (2012a): 42-48.
- "The pain of social disconnection: examining the shared neural underpinnings of physical and social pain". *Nature Reviews Neuroscience* 13 (2012b): 421-434.
- Feldman, Fred. *Pleasure and the Good*. New York: Oxford U. Press, 2004.
- What is this Thing Called Happiness?* New York: Oxford U. Press, 2010.
- Gettier, Edmund. "Is Knowledge Justified True Belief?". *Analysis* 23.6 (1963): 121-123.
- Glassen, P. "A Fallacy in Aristotle's Argument about the Good". *The Philosophical Quarterly* 7 (1957): 319-322.
- Goldman, Alvin. "Discrimination and Perceptual Knowledge". *Journal of Philosophy* 73.20 (1976): 771-791.
- Goldstein, Irwin. "Happiness: The Role of Non-Hedonic Criteria in Its Evaluation". *International Philosophical Quarterly* 13 (1973): 523-534.
- "Hedonic Pluralism". *Philosophical Studies* 48.1 (1985): 49-55.
- Hawkins, Jennifer. "Well-Being, Autonomy, and the Horizon Problem". *Utilitas* 20.2 (2008): 143-168.
- Haybron, Daniel. "On Being Happy or Unhappy". *Philosophy and Phenomenological Research* 71.2 (2005): 287-317.
- The Pursuit of Unhappiness*. New York, Oxford U. Press: 2008a.
- "Happiness, the Self and Human Flourishing". *Utilitas* 20.1(2008b): 21-49.
- Heathwood, Chris. "The Problem of Defective Desires". *Australian Journal of Philosophy* 83:4 (2005): 487-504.

- “The Reduction of Sensory Pleasure to Desire”. *Philosophical Studies* 133.1 (2007): 23-44.
- Hill, Scott. “Haybron on mood propensity and happiness”. *The Journal of Happiness Studies* 10.2 (2007): 215-228.
- Hurka, Thomas. *Perfectionism*. New York: Oxford U. Press, 1993.
- Kagan, Shelley. (1992). The limits of wellbeing. In E. F. Paul, F. D. Miller, & J. Paul (Eds.), *The good life and the human good*. Cambridge: Cambridge University Press.
- “Well-Being as Enjoying the Good”. *Philosophical Perspectives* 23 (2009): 253-272.
- Kitcher, Philip. “Essence and Perfection”. *Ethics* 110 (1999): 59-83.
- Kraut, Richard. “Two Conceptions of Happiness”. *The Philosophical Review* 88.2 (1979): 167-197.
- “Desire and the Human Good”. *Proceedings and Addresses of the American Philosophical Association* 68.2 (1994): 39-54.
- Lifton, Robert. *Thought Reform and the Psychology of Totalism*. Chappell Hill: U. of North Carolina Press, 1989.
- Keller, Simon. “Welfare as Success”. *Nous* 43.4 (2009): 656-683.
- McMahon, Darrin. *Happiness: A History*. New York: Atlantic Monthly Press, 2006.
- Montague, Roger. “Happiness”. *Proceedings of the Aristotelian Society* 67 (1966): 87-102.
- Morris, Stephen. “In Defense of the Hedonistic Account of Happiness”. *Philosophical Psychology* 24.2 (2011): 261-281.
- Nozick, Robert. *Philosophical Explanations*. Harvard U. Press, 1981.
- Rachels, Stuart. “Six Theses about Pleasure”. *Philosophical Perspectives* 18 (2004): 247-267.
- Raibley, Jason. “Happiness is not Well-being”. *Journal of Happiness Studies* 13.6 (2012): 1105-1129.
- Railton, Peter. “Facts and Values”. *Philosophical Topics* 14.2 (1986): 5-31.
- Rosati, Connie. “Persons, Perspectives, and Full-Information Accounts of the Good”. *Ethics* 105.1 (1995): 296-325.
- Russell, James. “Core Affect and the Psychological Construction of Emotion”. *Psychological Review* 110.1 (2003): 145-172.
- Tatarkiewicz, Wladyslaw. *Analysis of Happiness*. Trans. Edward Rothert and Danuta Zielinskn. Warsaw, Polish Scientific Publishers: 1976.
- “Sadness”. *The Concise Corsini Encyclopedia of Psychology and Behavioral Science*. Hoboken: Wiley, 2004. *Credo Reference*.
- Sidgwick, Henry. *The Methods of Ethics*. Indianapolis: Hackett, 1981.
- Singer, Margaret. *Cults in Our Midst*. San Francisco: Jossey-Bass Publishers, 1996.
- Skow, Raymond. “Preferentism and the Paradox of Desire”. *Journal of Ethics and Social Philosophy*. 3.3 (2009): 1-16.
- Smuts, Aaron. “The Feels Good Theory of Pleasure”. *Philosophical Studies* 155.2 (2011): 241-265.
- Sobel, David. “Full Information Accounts of Well-Being”. *Ethics* 104.4 (1994): 784-810.
- Sumner, Wayne. *Welfare, Happiness, and Ethics*. New York: Oxford U. Press, 1996.
- Schwitzgebel, E. “The unreliability of naive introspection”. *Philosophical Review* 117 (2008): 245-273.
- Vitrano, Christine. *The Nature and Value of Happiness*. Boulder, CO: Westview Press, 2013.
- Watson, David. “Positive Affectivity: The Disposition to Experience Pleasurable Emotional States”. *Handbook of Positive Psychology*. Eds. C.R. Snyder and Shane Lopez. New York: Oxford U. Press, 2002.
- Watson, David and Lee Anna Clark. “Measurement and Mismeasurement of Mood: Recurrent and Emergent issues”. *Journal of Personality Assessment* 68.2 (1997): 267-296.
- Whiting, Jennifer. “Aristotle’s Function Argument: A Defense”. *Ancient Philosophy* 8 (1988): 33-48.
- Wilkes, Katherine. “The Good Man and the Good for Man in Aristotle’s Ethics”. *Mind* 87 (1978): 553-571.
- Williams, Bernard. *Morality: An Introduction to Ethics*. Cambridge: Cambridge U. Press, 1976.
- Wilson, John. “Happiness”. *Analysis* 29.1 (1968): 13-21.

- Zablocki, Benjamin. "The Blacklisting of a Concept: The Strange History of the Brainwashing Conjecture in the Sociology of Religion". *Nova Religio: The Journal of Alternative and Emergent Religions* 1.1 (1997): 96-121.
- "Exit Cost Analysis: A New Approach to the Scientific Study of Brainwashing". *Nova Religio: The Journal of Alternative and Emergent Religions* 1.2 (1998): 216-249.
- "Reply to Bromley". *Nova Religio: The Journal of Alternative and Emergent Religions* 1.2 (1998): 267-271.
- "Towards a Demystified and Disinterested Scientific Theory of Brainwashing". *Misunderstanding Cults: Searching for Objectivity in a Controversial Field*. Eds. Benjamin Zablocki and Thomas Robbins. Buffalo: U. of Toronto Press, 2001. 159-214.
- Zimmerman, Michael. "Understanding What's Good For Us". *Ethical Theory and Moral Practice* 12 (2009): 429-439.