

A Defense of Basic Prudential Hedonism

by

Joe Nelson

Department of Philosophy
Duke University

Date: _____

Approved:

Walter Sinnott-Armstrong, Advisor

Jennifer Hawkins

Gopal Sreenivasan

David Wong

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Philosophy in the Graduate School
of Duke University

2020

ABSTRACT

A Defense of Basic Prudential Hedonism

by

Joe Nelson

Department of Philosophy
Duke University

Date: _____

Approved:

Walter Sinnott-Armstrong, Advisor

Jennifer Hawkins

Gopal Sreenivasan

David Wong

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Chemistry in the Graduate School of
Duke University

2020

Copyright by
Joe Nelson
2020

Abstract

Prudential hedonism is a school of thought in the philosophy of welfare that says that only pleasure is good for us in itself and only pain is bad for us in itself. This dissertation concerns an especially austere form of prudential hedonism: basic prudential hedonism (BPH). BPH claims that all pleasure is good for us in itself, and all pain is bad for us in itself, without exception; that all pleasures feel fundamentally alike, as do all pains; and that the amount of welfare in a person's life can (in principle) be calculated just by adding up the amount of pleasure it contains and subtracting the amount of pain. The dissertation presents a positive argument for the claim that pleasures and pains are defined by common phenomenal properties, defends BPH against a battery of objections, and outlines an argument for accepting BPH on the grounds that it is simpler than rival theories.

Contents

Abstract	iv
1. Introduction	1
1.1 Terminology	4
1.1.1 Evaluative Terminology.....	4
1.1.2 Hedonic Terminology	6
1.2 BPH and Its Competitors	10
1.2.1 The Fundamental Appeal of Basic Prudential Hedonism.....	10
1.2.2 Non-Hedonist Competitors	13
1.2.3 Hedonist Competitors	19
1.3 The Plan.....	23
2. Descriptive Objections.....	28
2.1 Phenomenological and Attitudinal Theories of Hedonic States	29
2.2 An Argument for the Phenomenological Thesis	32
2.2.1 The Argument	32
2.2.2 Non-Phenomenological Alternatives	35
2.2.3 Phenomenological Alternatives	47
2.3 General Arguments Against the Phenomenological Thesis	49

2.4 A General Argument Against the Calculation Thesis	58
2.5 Separate Experience or Hedonic Tone?	63
2.5.1 Idiosyncrasy Objections	63
2.5.2 Epistemic Objections	67
3. Evaluative Objections.....	72
3.1 Prudential Insight and the Undermining Strategy	73
3.1.1 Life Comparisons and the Strong Insight Hypothesis.....	73
3.1.2 The Undermining Strategy	82
3.2 Is the Undermining Strategy Self-Defeating?	93
3.3 More Evaluative Objections	99
3.3.1 Evaluative Arguments Against the Calculation Thesis.....	100
3.3.2 The Resonance Constraint	105
3.4 Conclusion	113
4. Simplicity	115
4.1 Syntactic Simplicity and Epistemic Probability	117
4.1.1 A Syntactic Criterion of Theoretical Simplicity	117
4.1.2 Epistemic Probability.....	122
4.2 The Simplicity of Basic Prudential Hedonism	128
4.2.1 Minimal Viability Conditions for Theories of Welfare.....	128

4.2.2 Simplicity Comparisons	132
4.3 Challenges.....	137
4.3.1 The Problem of Selecting Predicates	137
4.3.2 Universal Quantifiers and Skepticism	140
4.3.3 Is BPH Really Simpler?.....	141
4.4 Conclusion	144
5. Conclusion	145
References	147

1. Introduction

This dissertation defends a version of prudential hedonism. In all its forms, prudential hedonism says that only pleasure is intrinsically good for us and only pain is intrinsically bad for us.¹ Different versions of the view are distinguished by what they take pleasure and pain to be, and which pleasures and pains are said to count. But in every case, prudential hedonism makes the radical claim that nothing other than pleasure and pain—not love, friendship, autonomy, knowledge, achievement, failure, ignorance, helplessness, betrayal, or the satisfaction or frustration of desires—makes any difference, in itself, to the quality of a life for the person living it.

This claim has faced a number of influential criticisms. As a result, many philosophers regard prudential hedonism as untenable, and it is often described as unpopular. Yet it has always struck me as the most plausible approach in the philosophy of welfare. I began my research in this area with an inchoate sense that prudential hedonism was correct; this dissertation represents my effort to turn that basic impulse into something philosophically respectable.

Prudential hedonism, as I understand it, is not a single theory, but rather a family of theories, united by the radical claim stated above. Particular versions of prudential hedonism, when fully articulated, will specify what pleasure and pain are, which pleasures and pains count for and against welfare in themselves, and how pleasures and

¹ Whether “pleasure” and “pain” are the right terms to use to characterize prudential hedonism, or any other hedonistic theory, is disputed. See §1.1.2.

pains together determine the amount of welfare in a person's life. In principle, the number of such theories is vast, with each taking on some distinct metaphysical or evaluative commitment, and each allowing for different avenues of response to critics. Therefore, for practical reasons, I will not be attempting a defense of prudential hedonism *per se*.

Instead I will defend a single theory, which I call "basic prudential hedonism" (BPH). I believe that this theory presents the impulse behind prudential hedonism in its most primitive form. It is also an appropriate target for every canonical objection to prudential hedonism that I know of, in the sense that if any of those criticisms is well-founded, then BPH is almost certainly false. In this context, that is a virtue. By seeing what can be said on behalf of BPH, we can discover where prudential hedonism is more resilient against the canonical objections than commonly thought; by seeing where (if anywhere) it is indefensible, we will learn what sorts of concessions prudential hedonism will have to make to be a viable approach in the philosophy of welfare.

BPH says roughly this: all and only pleasure, broadly construed, is intrinsically good for you; all and only pain, broadly construed, is intrinsically bad for you; and how good your life is for you can (in principle) be calculated by adding up all of your pleasures and subtracting all of your pains. More precisely, we can understand it as consisting of the following theses, each of which contains a set of thematically linked claims. The first two theses are descriptive; they express BPH's central non-evaluative claims about pleasure and pain. The third thesis is evaluative; it expresses BPH's central claims about what makes our lives better or worse for us.

Basic Prudential Hedonism (BPH):

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

Calculation Thesis: All hedonic states have magnitudes that are

- (a) Determined solely by their intensities and durations.
- (b) Quantifiable (in principle, using cardinal numbers).
- (c) Commensurable with all others of their type (pleasures magnitudes with one another and pain magnitudes with one another).²

Evaluative Thesis: The amount of welfare in a life (or life-segment) L, for its subject S, is determined by, and equal to, the sum of the magnitudes of all of S's pleasures during L minus the sum of the magnitudes of all of S's pains during L.

I am partial to this theory. I believe that if any positive (i.e. non-skeptical) theory of welfare is correct, it is this one. But I do not believe this can be shown beyond reasonable doubt.

At various points, my case for BPH will depend on assumptions that I cannot fully defend here. I will try to be candid about what these assumptions are and where they come into play.

My defense of BPH is organized around two aims, one relatively modest, the other more ambitious. The relatively modest aim is to show that BPH is no less plausible than other theories of welfare. As far as this part of the dissertation is concerned, I would be satisfied if I could fight BPH's competitors, hedonist and non-hedonist alike, to a draw.

The more ambitious aim is to show that BPH has a distinct advantage over its major

² Notably, the calculation thesis does not claim that pleasure magnitudes are commensurable with pain magnitudes (though it does not deny this either). I omit this claim because I think that it would be superfluous. All that BPH needs is the claim, contained in the evaluative thesis, that amounts of intrinsic prudential goodness are commensurable with amounts of intrinsic prudential badness. On this point see Klockslem (2010).

competitors: its simplicity. But simplicity, as I understand it, can be no more than a tiebreaker in theory selection, so achieving the ambitious aim requires achieving the modest aim first. The idea is this: simpler theories are preferable, all else being epistemically equal; if I meet the modest aim, then I will have shown all else is indeed epistemically equal.

The purpose of this chapter is to lay the terminological, conceptual, and epistemic foundations for the work to follow. In §1.1, I will clarify my use of some key terms. In §1.2, I will make some remarks about the fundamental appeal of BPH before briefly introducing the reader to its major competitors in the philosophy of welfare, including some alternative formulations of prudential hedonism. Finally, in §1.3 I will briefly sketch my plan for meeting this work's modest and ambitious aims.

1.1 Terminology

1.1.1 Evaluative Terminology

This dissertation is a work in value theory. Accordingly, I will use the term “value” many times. When I use that term without qualification, I intend for it to cover both positive and negative value. In my parlance, to say that something is valuable is just to say that it is not evaluatively neutral. Something is valuable just in case it is good to some extent, bad to some extent, or both. I will use the phrases “positive value” and “negative value” for things that are (to some extent) good or bad, respectively.

Basic prudential hedonism is a positive theory of prudential value. Prudential value is the kind of value something has when it is good or bad for a person. In other words, it is the kind of value something has if it makes a contribution, positive or negative,

to an individual's welfare or well-being (I use the latter two terms interchangeably). In still other words, it is the kind of value something has when it is in a person's self-interest to have or avoid that thing. And in still other words, prudential value is the kind of value something has when it makes life better or worse for the one who has it. Positive theories of prudential value say that there are in fact things in the world with positive and/or negative prudential value. I will also refer to these as positive theories of welfare.

Like other kinds of value, prudential value can be intrinsic or extrinsic. Something has intrinsic value just in case it is valuable in itself, apart from anything else that it is related to. Something has extrinsic value just in case its value depends on its relations to other things. There are two main kinds of extrinsic value. The first is what I will call "attitude-dependent value". This is the kind of value something has when its value depends on its being the object of an attitude, such as a desire. The second, more important kind is instrumental value. This is the kind of value something has when, and because, it is conducive to getting things with intrinsic value.

Things with intrinsic prudential value are the fundamental building blocks of personal welfare. Everything that is good or bad for a person has this kind of value by being either intrinsically prudentially valuable, instrumentally prudentially valuable, or both. To illustrate: in this dissertation, I claim that pleasure has positive intrinsic prudential value; by extension, I claim that anything that brings a person pleasure has positive instrumental prudential value for that person. I find pleasure in lying down on my couch, so my couch has positive instrumental prudential value for me.

BPH is a type of hedonism. The qualifier “prudential” is required because the term “hedonism” is used for at least three different kinds of theory, all of which account for some area of philosophical interest entirely in terms of hedonic states. There is psychological hedonism, which says that all motivation can be explained in terms of hedonic states (for discussion see Sober 2013: pp. 131-140). There is ethical hedonism, which says that only hedonic states have intrinsic moral value (for example Lafleur 1956). And finally, there is prudential hedonism, also known as welfare hedonism, which says that only hedonic states have intrinsic prudential value. The present work is an exploration and defense of a form of prudential hedonism. It is not directly concerned with psychological or ethical hedonism.

1.1.2 Hedonic Terminology

As I have said, “hedonism” is a term for a type of theory that accounts for something philosophically interesting entirely in terms of hedonic states. To have a clear idea of what this means, we must have some idea of what hedonic states are. But it is challenging to give any definition of hedonic states without making some disputed assumption. A similar difficulty occurs at the level of terminology, as it becomes difficult to find terms for referring to hedonic states that fit well with ordinary usage and do not appear to unduly privilege some controversial conception of them.

With that in mind, I will begin with this very general characterization. A hedonic state is a type of mental state that essentially involves some phenomenology, or qualia. This is to say that every time a person is in a hedonic state, they must be feeling something;

there must be something that it is like for them to be in that particular token state.³ Moreover, every hedonic state has a valence, positive or negative. We commonly say of the positive hedonic states that they feel good and of the negative ones that they feel bad. If you have ever felt happy, sad, pleased, hurt, uncomfortable, giddy, or depressed, you have experienced a hedonic state.

I use “pleasure” for all positive hedonic states and “pain” for all negative ones, a practice I believe has a reasonably firm foundation in common usage. However, various philosophers who write on the value of hedonic states find that these terms have unduly narrow connotations. There is some merit to this. The word “pleasure”, in many contexts, calls to mind a restricted class of bodily pleasures—from food, sex, and the like—while pain, in common parlance and when discussed by philosophers and scientists, often refers to what we might call sensory pain: feelings that are experienced as having a more or less specific bodily locations, and that can be described with terms like burning, stinging, stabbing, aching, and throbbing.

I join the likes of Bentham (1988), Mill (2001), and Sidgwick (1981) in saying that while bodily pleasures and sensory pains are indeed cases of positive and negative hedonic states, there are also non-bodily pleasures and non-sensory pains.⁴ There is the

³ Even this assumption may be controversial, as it is incompatible with the attitudinal hedonism of Fred Feldman (2006), according to which hedonic states are propositional attitudes that do not necessarily have any felt quality. I will discuss aspects of Feldman’s view in chapters 2-4.

⁴ In defining the notion of utility, Bentham (1988: pp.14-15) writes:

By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered.

pleasure of finding something funny, for instance, and the pain of grief. But the more restrictive connotations that “pleasure” and “pain” often carry has led some philosophers to look for more suitable terms. Hence Mill (2001) uses “happiness” alongside “pleasure”; Crisp (2006) uses “enjoyment” and “suffering”; Schroeder (2004) uses “pleasure” and “displeasure”; Mayerfeld (1992) uses “happiness” and “suffering”; and Rachels (2004) prefers “unpleasure” as the antonym of “pleasure”.

Nonetheless, I maintain that “pleasure” and “pain” are the best options, as each of the foregoing alternatives has some disqualifying disadvantage.

The main alternatives on the negative side are “suffering”, “displeasure”, and “unpleasure”. “Suffering” is suitably broad, and neutral in itself as to the precise nature of the states it refers to. Unfortunately, “suffering” lacks grammatical versatility. “Pain” can take the form of a count noun, mass noun, adjective (“painful”), adverb (“painfully”), and verb (as in the phrase “it pains me”) without leaving the realm of ordinary speech. “Suffering”, by contrast, generally appears only as a verb (to suffer) or a gerund (as in the phrase “suffering is bad”). Adopting “suffering” as our general term for negative hedonic states threatens to require the use of phrases like “suffering-inducing” (compare

Mill (2001: p. 8) refers to the “pleasures of the intellect, of the feelings and imagination, and of the moral sentiments” alongside “those of mere sensation”. Sidgwick (1981, meanwhile, says various things indicating that he does not confine pleasure and pain to the physical realm. For example, in giving his account of what pleasure is, he writes: “when I reflect on the notion of pleasure—using the term in the comprehensive sense I have adopted, to include the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments...”. As for pain, he says that it “must be reckoned as the negative quantity of pleasure” (p. 124); given that pleasure can be intellectual, emotional, and sensual, it stands to reason that its “negative quantity” could be any of those things as well.

“painful”), “experience of suffering” (compare “painful experience”), and perhaps reference to individual “sufferings”. I prefer to avoid this kind of linguistic awkwardness.

“Displeasure”, meanwhile, seems suitable only in cases of relatively mild negative feeling. To describe, say, severe torture as “extremely displeasing” seems less like a literal truth and more like a bit of deliberately comic understatement (compare “extremely painful”). Moreover, both “displeasure” and “unpleasure”, as terms, suggest that they state the refer to is supposed to be the opposite of pleasure. But I wish to remain agnostic on this point. As far my defense of BPH is concerned, positive and negative hedonic states need only be opposed to one another at the evaluative level; it need not be the case that they are opposites at the descriptive level.

There is also the option of just using the phrases “positive hedonic state” and “negative hedonic state”. My objection to these is purely aesthetic: I find them too cumbersome to use as frequently as I would need to use them in this dissertation.

Admittedly, using “pleasure” and “pain” has its own disadvantages. Aside from being out of step with narrower usage elsewhere, it commits us to saying that experiences such as having a bad taste in your mouth, being tickled too much, and feeling bored are literally painful. But by now it should be clear that there is no terminology which will keep us from ever having to say anything that sounds clumsy or otherwise slightly odd.

Ultimately, the terminological differences between myself and other philosophers should not be very important. Those who believe a different set of terms is more faithful to the underlying facts may simply regard “pleasure” and “pain” as technical terms in what follows. So long as I am clear about what I mean by “pleasure”, “pain”, and related

terms, it will be possible to translate every statement in this dissertation that includes one or more of those words into the preferred terminology of any of the philosophers mentioned above. For example, a philosopher who believes that “displeasure” is the proper umbrella term for negative hedonic states may simply replace all my uses of “pain” with “displeasure”, and be confident that no loss of meaning will result.

1.2 BPH and Its Competitors

1.2.1 The Fundamental Appeal of Basic Prudential Hedonism

I turn now to the topic of what makes basic prudential hedonism an appealing view. What follows here is not an argument for BPH. Rather, it is a short exploration of what might draw someone to it in the first place.

It is common for works on the philosophy of well-being to begin by discussing some form of prudential hedonism. Typically, this is done to show how prudential hedonism is mistaken, in order to motivate whatever theory of well-being the author ultimately wants to defend.⁵ Still, the fact that hedonism is such a common starting place for explorations of prudential value is striking. It suggests that, for all its unpopularity, something about this approach is compelling. This owes, I suspect, to the character of pleasurable and painful experiences. There is something about the way pleasure feels that makes us think that it’s good for us, and something about the way that pain feels that

⁵ Examples include Griffin (1986: pp. 7-9), Kagan (1994: pp. 310-311), Hurka (2011: pp. 65-73), and Fletcher (2016: pp. 14-24). Despite its already poor reputation, philosophers still sometimes publish articles devoted entirely to criticizing hedonism. See for example Sobel (2002) and Dorsey (2011).

makes us think that it is bad for us. Evaluative language attaches to the experiences so naturally that it seems redundant to say that pleasure feels good and pain feels bad.⁶

Furthermore, prudential hedonism fits with the commonsense idea that we are at least sometimes motivated to pursue things that are good for us. Virtually everyone generally seeks pleasure and avoids pain. And when we do forego pleasure or choose to endure pain, it is almost always in the service of something else that we value. This something else is often more pleasure and/or less pain in the future.

Granted, this does not show that pleasure and pain have intrinsic prudential value. And even if it did, one need not be a prudential hedonist to believe that pleasure and pain are good and bad for us in themselves, respectively. One could claim that hedonic states have intrinsic prudential value and other things do too. As we will see in §1.2.2 below, there are philosophers who hold views like this. Among the things other than pleasure that philosophers of welfare have proposed are intrinsically good for us are friendship, achievement, autonomy, knowledge (of an important or at least non-trivial sort), and moral virtue.

But notice that, for most people under normal circumstances, these very things tend to bring us pleasure, help us avoid pain, or both. Not all the time, of course; gaining and preserving them can involve foregoing pleasure and enduring considerable suffering. But spending time with friends and accomplishing significant things are among life's great sources of joy. And getting to experience them with any consistency typically

⁶ Mendola (1990) develops this insight into an argument for the objective value of pleasure and pain.

requires some measure of kindness, honesty, patience, courage, and the like, as well some non-trivial knowledge and some freedom to choose one's own course of life. There is also the fact of how painful it is (for most people, under normal circumstances) to go without such things. We seek close relationships with others, and try to do worthwhile things with our lives, in no small part to avoid the suffering we would experience if we did not. This is one reason why the less pleasant feelings involved in these pursuits are often regarded as "worth it", so to speak.

Prudential hedonism takes these observations and fashions an entire theory out of them. It claims that the prudential value of any given thing that is not a hedonic state can be accounted for entirely in terms of the conduciveness of that thing to pleasures, pains, and/or absences thereof. In other words, it says that something like friendship is good for us not in itself, but because it makes our lives more pleasurable and/or less painful than they would be otherwise.

Now, a prudential hedonist need not say that all pleasures are good for us, or all pains bad for us. They may claim that only a subset of pleasures are good or pains bad. But BPH does affirm the intrinsic prudential goodness of all pleasures and the badness of all pains, without exception. Hence BPH proposes that the fundamental facts about welfare can be just this simple: the good feelings that we generally seek are in fact what's intrinsically good for us, the bad feelings we generally avoid are in fact what's intrinsically bad for us, and that's all there is to it. Further, it proposes that how good or bad something is for you is just a matter of adding up the amount of pleasure it brings you and subtracting the amount of pain it brings you.

All in all, then, BPH paints a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. It assigns intrinsic prudential value to things that it is *prima facie* plausible to think have an impact on our welfare, thereby vindicating the commonsense idea that it is generally to a person's benefit to have mutually rewarding relationships, success at worthwhile activities, strong moral character, and freedom from sickness, injury, and subjugation by others. And it does so with what appears to be relatively few theoretical trappings.

Implicit in BPH is a challenge: why make your theory of welfare any more complicated than this? Philosophers have answered that challenge in many ways. Hence this dissertation.

1.2.2 Non-Hedonist Competitors

Following Parfit's (1984: pp. 493-502) influential discussion, the standard taxonomy of theories of welfare has consisted of three categories: hedonism, desire theories, and objective list theories. I will not depart from this standard. In §1.2.3. I will describe the main non-BPH variants of prudential hedonism. Here, I will give a brief summary of the desire and objective list approaches.

Desire theories of welfare hold that only the satisfactions and frustrations of desires have intrinsic prudential value. These are not to be confused with feelings of satisfaction or frustration. A desire is satisfied when the desired state of affairs occurs, and frustrated when the desired state of affairs does not occur. Every desire theory that I know of holds that only the satisfactions or frustrations of a person's own desires can have intrinsic prudential value for that person. There is, however, a great deal of variation

among extant desire theories on other matters, such as which desires count, and whether the desires that fundamentally matter are actual or hypothetical.

Actualist desire theories allow only the satisfactions and frustrations of a subject's actual desires to count. Among actualist theories, there are restricted and unrestricted views.

Unrestricted desire theories say that all of a person's desire-satisfactions and desire-frustrations have intrinsic prudential value (Murphy 1999, Lukas 2010). So, take any desire of mine, be it as mundane as a desire for pleasure or as eccentric as a desire about the exact number of hairs on another person's head; unrestricted desire theories hold that if this desire is satisfied, my welfare goes up, and if it is frustrated, my welfare goes down.

Some unrestricted actualist desire theories fall prey to a paradox. The best-known exposition of this comes from Bradley (2007), though he credits Feldman (2006: pp. 16-17) with being the first to notice it. The problem is that if we accept certain common assumptions about desires, and do not place any restrictions on which satisfactions and frustrations of desires can have intrinsic prudential value, we can create a logically possible situation in which a person's life contains an overall positive and negative amount of welfare simultaneously.

Here's an example of a case that produces the paradox. Suppose we can assign numerical strengths to desires. Further suppose that if I have a desire of strength X , then satisfying that desire is intrinsically positively valuable for me to degree X , and frustrating that desire is intrinsically negatively valuable for me to degree X . Now, imagine a person

who has exactly two desires: a desire of strength 1 not to feel pain, and a desire of strength 2 that their life contain an overall negative amount of welfare. If this person starts at a welfare level of 0, and then feels some pain, their desire not to feel pain will be frustrated, and their welfare will sink to -1. But then having a negative welfare level will satisfy their other desire, bringing them to a welfare level of +1. This frustrates their desire to have a negative welfare level, which brings them back down to -1. And so on.⁷

This and other problems with unrestricted actualism provide some motivation for moving to a restricted actualist theory.⁸ Restricted actualist views say that only a subset of a person's actual desire-satisfactions and desire-frustrations have intrinsic prudential value. For instance, one might say that only intrinsic desires—desires directed at things for their own sake, rather than merely as a means to something else—are fundamentally relevant. Most restricted views go further than this: for example, Griffin (1986) argues that only one's properly informed actual intrinsic desires count; Heathwood (2006) suggests that only actual intrinsic desires about one's present state of consciousness count; and Heathwood (2019) argues that only the actual intrinsic desires that count as "genuine attractions", as opposed to mere behavioral dispositions, count.⁹

⁷ Skow (2009) argues that if desire-satisfaction and -frustration comes in degrees, rather than being all-or-nothing, then a desire theory can be developed that avoids the paradox. I think he is mistaken. Even if desire-satisfactions and -frustrations come in degrees, it is still conceptually possible to desire that your life be overall bad for you while being otherwise indifferent to how bad it is. Such a desire would be fully satisfied when your welfare is below zero to any extent, and fully frustrated when your welfare is above zero to any extent, thereby producing the paradox.

⁸ The Bradley/Feldman paradox notwithstanding, the most prominent objection to unrestricted actualism is that some desires are just too trivial for their satisfactions or frustrations to have intrinsic prudential value. Lukas (2010) defends unrestricted actualism against this objection.

⁹ The restriction to informed desires (e.g. Griffin 1986) avoids the paradox on the plausible assumption that the desire that one's life go poorly would not be an informed one. The restriction

Ideal desire theories, meanwhile, assign intrinsic prudential only to the satisfactions and frustrations of the desires that a person would have under certain ideal conditions. Though he speaks in terms of objective interests and what is non-morally good for a person, rather than using the language of welfare or prudential value, we might understand Railton (1986a: pp. 176-184) as presenting a view of this type. He proposes that what is good for a person (call him J) is to get what an ideally informed and rational version of that person (call him J+) would want for the actual person to have. Here is an illustration: J is very thirsty, and has the choice to drink either water or wine. He wants to drink the wine, and for some odd reason, has no desire to drink the water. But unbeknownst to him, he has suddenly developed an allergy to alcohol, and so the wine will make him sick. J+, however, is ideally informed, and therefore knows about J's allergy. Hence J+ wants J to want to drink the water. As a result, though drinking the wine would satisfy one of J's actual desires, drinking the water is what would actually be good for him.

Like BPH, desire theories can paint a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. There is a close intuitive connection between well-being and getting what you want (and avoiding what you don't want). Moreover, most of us want to be morally good people, and to have friendship, achievement, autonomy, and at least some kinds of knowledge in our lives. So

to intrinsic desires about present states of consciousness (e.g. Heathwood 2006) avoids it on the plausible assumption that the desire for one's life to go poorly overall is not such a desire.

just as having these things is associated with getting more pleasure and less pain, they are also associated with having more desires satisfied and fewer frustrated. And it is reasonable to think that on a restricted desire theory like those cited above, desires for friendship, achievement, autonomy, and at least some kinds of knowledge would be among those whose satisfactions and frustrations have intrinsic prudential value (in other words, these desires would not be filtered out by the theory's restrictions). Hence desire theorists can also vindicate the commonsense idea that it is generally to a person's benefit to have mutually rewarding relationships, success at worthwhile activities, strong moral character, and freedom from sickness, injury, and subjugation by others.

Another major motivation for desire theory as an approach in the philosophy of welfare is that it allows for a close connection between one's welfare and one's particular psychology. The less restrictive the desire theory, and the more closely tied it is actual desires, the less it dictates to each person what they must do for their life to be good for them. Even Railton, whose theory abstracts away from actual desires the most of any considered here, holds that "what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware", adding that "it would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him" (1986b: p. 9).¹⁰

¹⁰ I defend BPH against the claim that there is a necessary connection between intrinsic value and what a person would find compelling or attractive at §3.3.3.

However, in the progression from unrestricted actualism to restricted actualism to ideal desire theory, we can trace an increasing resistance to the notion that what is good for a person must be in some way sensitive to individual idiosyncrasies. This brings us, finally, to a type of view that dispenses with any such requirement: objective list theory.¹¹

Objective list theories of welfare posit multiple attitude-independent sources of intrinsic prudential value. In other words, they provide a list of things that are good for you whether you, or any hypothetical version of you, has any positive attitude toward them or not.¹² For example, Fletcher (2013) proposes that the list of objective prudential goods consists of pleasure, friendship, achievement, happiness, self-respect, and virtue, while Hooker (2015) argues that the list consists of pleasure, friendship, significant achievement, important knowledge, and autonomy. (Notice that objective list theorists can, and often do, assign positive intrinsic prudential value to pleasure.) We may also classify the perfectionist approach to welfare as a kind of objective list theory. Perfectionists, such as Hurka (1993) and Kraut (2007), locate intrinsic prudential value in

¹¹ Objective list theories need not hold that what has intrinsic prudential value for a person is unconnected to their attitudes. They may say that some items on the list are partly constituted by certain attitudes; for instance, they could say that friendship is partly constituted by attitudes that friends hold toward one another. But they need not say any such thing. And in all cases, objective list theorists hold that the items on the list have intrinsic prudential value regardless of your attitude toward them. So even if friendship, for instance, is partly constituted by certain attitudes, you need not have any attitude toward friendship itself in order for it to be intrinsically good for you.

¹² For reasons that are unclear to me, objective list theorists do not seem much interested in negative prudential value. Gert (2005: pp. 90-99) is an exception; he himself notes that “[i]n most discussions of goods and evils, goods receive most of the attention. Indeed, sometimes evils are completely ignored, almost as if they did not exist” (p. 90).

the development and expression of distinctively human capacities, such as the pursuit of knowledge and the exercise of moral virtue.

Like prudential hedonists and desire theorists, objective list theorists can paint a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. For this, they need only be judicious in what they put on their lists. And if they assign positive and negative intrinsic value to pleasure and pain (respectively), then they can avail themselves of many of the observations I appealed to on behalf of prudential hedonism above. Hence they can easily vindicate the commonsense idea that it is generally to a person's benefit to have mutually rewarding relationships, success at worthwhile activities, strong moral character, and freedom from sickness, injury, and subjugation by others.

The taxonomy I have given in this section is certainly not exhaustive, but it gives us a good enough sense of the terrain in the philosophy of welfare. Ultimately, the finer points of these competing views are less important than the general properties that distinguish them from BPH. It is with reference to these properties that I will discuss the simplicity of BPH in chapter 4.

1.2.3 Hedonist Competitors

I know of only one philosopher who explicitly endorses basic prudential hedonism (albeit not under that name): Tännsjö (1998). The view is sometimes attributed to Bentham (1989), but I'm not sure this is right. Bentham clearly accepted the evaluative thesis, and may have accepted the phenomenological thesis. But in his account of how to calculate the goodness of a pleasure for an individual, Bentham cited not only the intensity and

duration of the pleasure as factors, but also its “certainty or uncertainty” and its “nearness or remoteness” (p. 29). I do not know whether Bentham meant for these to figure into the intrinsic goodness of the pleasure of the individual or not. In any case, neither Bentham nor Tännsjö defend their versions of hedonism at great length. Other than this dissertation, I have not personally encountered any extended defenses of BPH in the literature.

Every other form of prudential hedonism can be understood as deviating from BPH by replacing one of its three theses with something more complicated. This is what makes BPH “basic”, and a natural starting point for assessing the strengths of hedonism as an approach in the philosophy of welfare. When I say that other kinds of hedonism are more complicated than BPH, I mean roughly that they must make a greater number of fundamental claims. I will be more precise about this in chapter 4. For now, I will have to trust that the reader shares my general sense of when a theory must make an additional fundamental claim.

First, consider BPH’s phenomenological thesis. This thesis holds that all pleasures feel alike in some particular way, as do all pains, and that it is these phenomenological similarities that define pleasure and pain as categories. Some forms of hedonism complicate this by adopting an attitudinal theory of hedonic states instead. Attitudinal theories say that each token hedonic state involves both a state of affairs (such as a feeling) and an attitude toward that state of affairs (such as a desire that it occur or not occur).

Feldman (2006: pp. 55-60), a contemporary prudential hedonist, holds an attitudinal view of hedonic states. According to Feldman, pleasures and pains are

propositional attitudes: to have pleasure is to have the propositional attitude of being pleased that some proposition is true, and to have pain is to have the propositional attitude of being pained that some proposition is true. Hence where BPH says that token pleasures and pains are individual mental states, an attitudinal view like Feldman's holds that token pleasures and pains necessarily involve pairs of states, consisting of attitudes and the states of affairs toward which those attitudes are held. In this respect, attitudinal versions of hedonism must be more complicated than BPH.

Then there are the prudential hedonists who accept the phenomenological thesis, but would replace the evaluative thesis with something more complicated. On my reading of him, Mill (2001: pp. 7-11) is one of these. Mill says that there are distinctions of quality among pleasures, such that the pleasures that appeal to the "higher faculties" of human beings—pleasures of the intellect, imagination, and moral character—have greater intrinsic value than the base pleasures of which humans and animals alike are capable. He appears to intend this distinction as being one purely of value, not of quantity; the higher pleasures are not literally more pleasurable for us, but they are nonetheless intrinsically better for us.¹³ This means that to determine the amount of welfare in a life, we need to do more than add up all the pleasure magnitudes and subtract all the pain magnitudes, as the evaluative thesis claims. To get the right evaluative result, Mill would

¹³ Passages indicating that Mill (2001: p.9, p. 10) thinks of the quality distinction as evaluative rather than as purely descriptive include his remark that "[i]t is quite compatible with the principle of utility to recognize the fact that some kinds of pleasures are more desirable and *more valuable* than others", and his observation later in the discussion of higher pleasures that "[m]en often, from infirmity of character, make their election for the nearer good, though they know it to be *less valuable*" (emphases mine).

have us perform a more complex calculation of value that accounts for quality distinctions among pleasures.

Another contemporary prudential hedonist, Crisp (2006: pp. 112-116), adopts Millian quality distinctions at the descriptive level. We can interpret him as accepting the phenomenological and evaluative theses, but rejecting the calculation thesis. For Crisp, as for Mill, there are higher and lower pleasures, and the higher pleasures are more valuable. But, Crisp says, they are more valuable precisely because they are more pleasurable (Crisp would say “enjoyable”). Therefore, in determining the magnitude of a pleasure, we must account for its intensity, duration, and quality. And disparities in quality can be such that no amount of a lower pleasure could ever be as pleasurable as any amount of a higher one: to take Crisp’s example, it may be that no amount of pleasure one gets from drinking lemonade could ever exceed the amount of pleasure one gets from reading *Pride and Prejudice*, even keeping the durations of the pleasures constant.¹⁴

Finally, Bramble (2016: pp. 98-101) also rejects the evaluative thesis, on the grounds that “purely repeated” pleasures do not contribute to a person’s total lifetime welfare. On Bramble’s view, if a person feels two separate pleasures of the same kind—two pleasurable bites of the same dessert, say—the second can only add to the person’s welfare if it somehow feels pleasurable in a different way than the first. This implies that the amount of welfare in a life is not determined by adding up all the pleasures and subtracting all the pains; we would also need to subtract the magnitudes of the pleasures

¹⁴ Unlike Mill, Crisp allows for quality distinctions among pleasures to vary from one person to another. A higher pleasure for you might be a lower pleasure for me.

that did not “introduce anything qualitatively new in terms of pleasurable-ness” to the subject’s life (p. 98).

Bramble adopts this view for much the same reason that both Mill and Crisp posited quality distinctions among pleasures: to avoid the implication that a life of nothing but base bodily pleasures could be as high in welfare, or higher, than a life with pleasures that are more varied and refined.¹⁵ The evaluative thesis ensures that BPH cannot avoid this implication.

1.3 The Plan

Chapters 2 and 3 are devoted to the relatively modest goal of the dissertation: showing that basic prudential hedonism is at least as defensible as other major positive theories of welfare, i.e. that we do not have adequate reason to reject BPH in favor of any of the theories outlined above. Chapter 4 is devoted to the more ambitious goal, which is to show that we have rationally ought to accept BPH over those other views. Throughout, I take for granted that skeptical positions in the relevant areas—the philosophies of hedonic states and welfare—are false. In other words, I assume that pleasure and pain exist; that some things really are intrinsically good or bad for us; and that we can, and do, know about these things. The structure of these chapters is as follows.

¹⁵ Bramble (2016: p. 98) asserts that “[a] life of purely bodily pleasures, unlike one involving some of the pleasures of love, learning, aesthetic appreciation, etc., can involve very little qualitative diversity in pleasures. Its pleasures quickly become ‘just more of the same’”.

In chapter 2, I discuss descriptive objections to BPH, i.e. objections relying on only non-evaluative premises. These objections attack BPH's descriptive theses, which are the phenomenological thesis and the calculation thesis:

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

Calculation Thesis: All hedonic states have magnitudes that are

- (a) Determined solely by their intensities and durations.
- (b) Quantifiable (in principle, using cardinal numbers).
- (c) Commensurable with all others of their type (pleasures magnitudes with one another and pain magnitudes with one another).

My defense against descriptive objections consists of four parts.

First, I provide some relevant background in competing theories of hedonic states. Here I introduce two views of pleasure and pain that are compatible with BPH: the separate experience view, on which pleasure and pain are each single feelings that could, in principle, be felt in isolation from anything else; and the hedonic tone view, on which pleasure and pain occur when qualia take on a certain distinctive phenomenal property that cannot be felt on its own. I also discuss attitudinal views of pleasure and pain, which reject the phenomenological thesis. These views have it that each instance of a hedonic state necessarily involves an attitude (such as a desire) directed at some state of affairs (such as an occurrent feeling).

Second, I present a positive case for the phenomenological thesis, in which I contend that this thesis best explains our ability to recognize when we are in token states of pleasure or pain. I argue that if we accept the phenomenological thesis, we can account for this ability easily, whereas no non-phenomenological approach is able to account for

it in a plausible manner. I grant that there are alternative phenomenological theories of hedonic states that reject BPH's phenomenological thesis, and that these can account for the explanandum in question. I reject these on the grounds that they are more complicated without have any distinct advantages.

Third, I respond to general arguments against the descriptive theses; that is, arguments that target the theses directly, rather than a separate-experience or hedonic tone view that is consistent with them. I show that these arguments rely on appeals to phenomenal introspection, and therefore on the assumption that this faculty is suitably reliable. I then consider and reject some rationales for accepting this assumption.

Fourth, I argue that although we should embrace the phenomenological thesis, we should remain agnostic between separate-experience and hedonic tone theories of pleasure and pain. Here I consider some arguments against each of these views, and show that none gives us adequate reason to favor one approach over the other.

In chapter 3, I respond to evaluative objections to BPH, i.e. objections that rely on one or more evaluative premises. Most of these take aim squarely at the evaluative thesis:

Evaluative Thesis: The amount of welfare in a life (or life-segment) L, for its subject S, is determined by, and equal to, the sum of the magnitudes of all of S's pleasures during L minus the sum of the magnitudes of all of S's pains during L

My defense against evaluative objections consists of three parts.

First, I rebut some arguments to the effect that things other than pleasure and pain have intrinsic prudential value. Here I make use of a defensive strategy I call the "undermining strategy", and contrast my version of it with similar defenses given by other prudential hedonists. Second, I respond to the objection that the undermining

strategy, as employed in defense of BPH, is epistemically self-defeating. Third, I apply the undermining strategy to three evaluative arguments that, in one way or another, concern BPH's descriptive theses: two against the calculation thesis, and one argument to the effect that if the phenomenological thesis is true, then the evaluative thesis is implausible. The latter argument claims that BPH fails to meet a (purported) constraint on theories of welfare that is sometimes called the "resonance constraint"; I conclude the chapter with a positive evaluative argument against this constraint.

In chapter 4, I move on to the ambitious aim: to show that BPH is the best positive theory of welfare available. My case for this centers on a theoretical virtue, much discussed in the philosophy of science but rarely, if ever, in value theory: simplicity. The kind of simplicity I have in mind is a kind of syntactic simplicity, where the simpler of two theories is the one that makes fewer fundamental claims. After defining this notion, I show that, given some assumptions about the epistemic probabilities of these fundamental claims, a theory's being simpler by this standard is, in a literal sense, more likely to be true. Finally, I argue that by this same standard, BPH is the simplest minimally viable theory of welfare.

Admittedly, I do not give any detailed defense of the assumptions that this argument relies on. Doing so would take us far afield from value theory into fundamental issues in epistemology and metaphysics. To the extent that this argument makes a significant contribution, then, it is not in solving any fundamental problems about theoretical simplicity or the epistemic value thereof; rather, it is in showing how an important question in value theory—that of how to decide between competing views of

welfare that have equal claims to fundamental appeal and resilience against standard objections—depends on whether, and how, we can make progress on deep questions in other areas of philosophy.

2. Descriptive Objections

Objections to basic prudential hedonism (BPH) come in two varieties: descriptive and evaluative. This chapter concerns the descriptive sort, which imply that BPH conflicts with non-evaluative facts about pleasure and pain. In effect, these objections reject one or both of BPH's descriptive theses:

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

Calculation Thesis: All hedonic states have magnitudes that are

- (a) Determined solely by their intensities and durations.
- (b) Quantifiable (in principle, using cardinal numbers).
- (c) Commensurable with all others of their type (pleasures magnitudes with one another and pain magnitudes with one another).

Of course, I cannot address every extant argument against these theses. I can only assure the reader that I have chosen representative examples of the most prominent argumentative strategies against these theses that I have yet encountered.

This chapter consists of five main parts. In §2.1, I set out two distinctions that are crucial to all that follows: between phenomenological and attitudinal theories of hedonic states, and between two sub-types of phenomenological theory that I call the "separate experience view" and the "hedonic tone view". Then, before playing defense, I spend some time on offense in §2.2, where I present a positive argument for the phenomenological thesis.

In §2.3, I respond to the most common sort of general challenge to the phenomenological thesis: that when we introspect, we do not detect any common phenomenal quality among various pleasures or among various pains. Here I argue that

objections of this kind rely on an unsupported claim about our introspective capacities, and are therefore unsuccessful.

In §2.4, I respond to a general argument against the calculation thesis. This argument is similar to those considered in the previous section, and accordingly, my reply proceeds along similar lines.

Finally, in §2.5, I address the question of whether we should adopt the separate experience view, the hedonic tone view, or remain agnostic between them. I argue for the last of these options: suspension of judgment as to which theory is correct. I survey representative arguments against each of these views, and explain why I think they fail and why their failure suggests that suspension of judgment is the appropriate stance.

2.1 Phenomenological and Attitudinal Theories of Hedonic States

Understanding the arguments of this chapter requires some background in the philosophy of hedonic states. This section is devoted to providing this background. Philosophical accounts of pleasure and pain fall mainly into two categories: phenomenological theories, which typically embrace the phenomenological thesis, and attitudinal theories, which do not. For the sake of clarity, I will explain the phenomenological/attitudinal distinction in terms of pleasure, with the understanding that these theories will say the same, *mutatis mutandis*, about pain.

Phenomenological theories of pleasure standardly hold that pleasures are unified by a common phenomenology (though see §2.2.2 below). On this sort of view, all pleasures, phenomenologically diverse though they may be, feel exactly the same way in some respect, and it is this common feeling that makes them all count as pleasures.

Some phenomenological theories claim that pleasure is a single feeling, separate from other qualia, such that it is at least logically possible for one to feel pleasure itself and nothing else (Bramble 2013). On this kind of theory, when one finds pleasure in (for example) a physical sensation, one's experience contains the feeling of pleasure itself alongside that sensation, with the two qualia overlapping in consciousness. I will refer to these as "separate experience" theories.¹⁶

Other phenomenological theories deny that pleasure can be experienced by itself, claiming instead that pleasure occurs when one or more of a subject's qualia take on a positive "hedonic tone" (Broad 1930: pp. 229-230). Dimensionalist theories, which can be understood as a subset of hedonic tone theories, describe pleasure as a phenomenological dimension along which experiences vary, analogous to other qualitative dimensions such as loudness. Just as a single noise can sound more or less loud, any given qualia can be more or less pleasurable; and just as there can be no qualitative loudness without sound, there can be no pleasure without some qualia to occupy a place along the pleasure dimension (Kagan 1992, pp. 172-175; Moen 2013).

Finally, some phenomenological theorists decline to take sides on the issue of whether pleasure can be felt in phenomenological isolation, preferring to simply assert that all pleasures feel alike in that they feel enjoyable (Crisp 2006: p. 109) or feel good (Smuts 2010).

¹⁶ This name is inspired by Moen (2013), who refers to this sort of view as "split experience" theory. I have opted for the word "separate" rather than "split" to more clearly convey the idea that pleasure and pain can be felt separately; on this view, there is nothing "split" about pleasure and pain themselves.

Attitudinal theories, by contrast, deny that there is phenomenological unity among pleasures. Instead, they claim that a subject's pleasures are unified by a common pro-attitude that the subject holds toward (some of) their own states of consciousness. For example, an attitudinal theorist might hold that a state of consciousness counts as a pleasure if and only if the subject intrinsically desires for it to be happening while it is happening; that is, desires for it to be happening for its own sake, not merely for the sake of some further end.

Heathwood (2007) defends an attitudinal view of sensory pleasure. Extended to pleasure in general, his theory provides a useful example of the attitudinal approach to hedonic states:

A feeling *F*, occurring at time *t*, is a pleasure at *t* iff the subject of *F* desires, intrinsically and *de re*, at *t*, of *F*, that it be occurring at *t*.¹⁷

Candidates for the attitude that unifies pleasure as a category include not only desire (see also Alston 1967), but also liking (Katz 1986, Brax 2009) and a *sui generis* propositional attitude of being pleased (Feldman 2006).

With all this in mind, I will now argue that the phenomenological thesis is right, and therefore that all attitudinal theories of hedonic states are mistaken. But I will not be taking sides between the separate experience and hedonic tone views. For reasons that will become clear, I do not think that we are justified in preferring either of these theories over the other.

¹⁷ This is adapted from Heathwood's RSPD3, which reads: "A sensation *S*, occurring at time *t*, is a sensory pleasure at *t* iff the subject of *S* desires, intrinsically and *de re*, at *t*, of *S*, that it be occurring at *t*" (2007: p. 32). Heathwood endorses the idea that his view could be extended to non-sensory pleasures (p. 28).

2.2 An Argument for the Phenomenological Thesis

2.2.1 The Argument

Once again, the phenomenological thesis is this:

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

My positive argument for this thesis involves an inference to the best explanation. The fact to be explained, which I will just call “the explanandum”, is this:

Explanandum: We sometimes become aware, at least partly through phenomenology, that we are in a state of pleasure or pain.

I will attempt to show that the best explanation of this fact presupposes that the phenomenological thesis is true.

Before I go on, four clarifications are in order. First, the word “we” in the above statement of the explanandum refers to we human beings. Second, the statement should be read as saying that we sometimes become aware that we are experiencing pleasure, and also that we sometimes become aware that we are experiencing pain—not just that we sometimes become aware of being in some kind of hedonic state. Third, as I use the term here, to be aware that X is to consciously know that X. So the statement asserts that we sometimes consciously know that we are in a state of pleasure or pain. Finally, to become aware of X “at least partly through phenomenology” is to become aware of X at least partly by having one or more mental states that, in themselves, affect what it is like to be you at the moment. More concisely (and colloquially), to become aware of X “at least partly through phenomenology” is to become aware of X at least partly by having feelings.

In summary, the explanandum says that in at least some cases, feeling things is part of the process by which you become aware of being in a state of pleasure or pain. It does not say that feeling things is always part of this process. Nor does it say that having a feeling is ever a necessary part of this process. Nor does it say that any specific feeling must be involved. It only says that some feeling is part of the process on some occasions.

I believe that this explanandum is an obvious fact. But even so, I will say two more things on its behalf.

In support of the point that we sometimes know that we are in a state of pleasure or pain, I will note that to deny this is to give up on theorizing about hedonic states altogether. If we do not ever know when we are in a state of pleasure or pain, then we have no way of telling whether our theories of pleasure and pain in any way match the phenomena they purport to be about. Hence this dissertation must presuppose that part of the explanandum, and indeed all the criticisms of BPH that I will discuss below must presuppose it as well. Consider also that to justify rejecting this part of the explanandum, we would need a rationally convincing argument implying (for example) that when a cognitively competent adult burns himself with a branding iron, attends to the subsequent feeling, and classifies it as painful, he does not actually know that he is in pain. I suspect that no such argument exists or is forthcoming.

In support of the point that knowledge of our hedonic states is at least sometimes conscious, and gained at least in part through phenomenology, I invite the reader to perform an experiment. Pinch yourself on the arm, increasing the pressure until you are aware of being in a state of mild pain. How did you become aware of when the pain

began? I would wager that you did so at least in part through phenomenology. During the experiment, you underwent changes in what it was like to be you at the moment—you felt things—and this was at least part of how you became aware of the onset of a pain. This shows that awareness of pain does sometimes occur at least partly via phenomenology.

Unfortunately, it's generally easier to gradually induce pain in ourselves than pleasure. But if you can, try to repeat the experiment by giving yourself a gentle scalp massage, starting with pressure light enough as to be nearly imperceptible, and gradually applying more until it feels pleasant. If that does not work, feel free to substitute any suitable pleasure-inducing activity. In any case, I expect you will find that detecting the onset of pleasure involved feeling something. This shows that awareness of pleasure, too, sometimes occurs at least partly via phenomenology.

If we suppose that the phenomenological thesis is true, these facts are easy to explain. We can become aware of our pleasures and pains through phenomenology because pleasures and pains are phenomenological states. Phenomenology provides the most direct path to awareness (conscious knowledge) of the properties that make our pleasures and pains what they are; the situation is analogous to using vision to become aware of visual qualia or hearing to become aware of auditory qualia. It's simply the right tool for the job.

So, the phenomenological thesis explains the explanandum. But this is not enough. For the phenomenological thesis to provide single best explanation, there must be adequate reason to favor this explanation over the alternatives. I cannot do the whole job

of giving such reason here; demonstration of some virtues of the phenomenological thesis—its resilience to standard objections, and its simplicity compared to certain rivals—will have to wait for later parts of the dissertation. But I will build a presumptive case for the phenomenological thesis by attempting to show that every non-phenomenological alternative is seriously flawed. I will also consider different sort of phenomenological view, and argue that it enjoys no distinct advantage over BPH's phenomenological theiss.

2.2.2 Non-Phenomenological Alternatives

I begin with accounts of the explanandum that presuppose some attitudinal theories of hedonic states. The basic structure of my reasoning with respect to these theories will then carry over to all other non-phenomenological accounts.

Recall that for attitudinal theorists, each token hedonic state involves both an attitude, such as a desire, and a state of affairs toward which that attitude is held, such as a feeling. For instance, a fairly crude attitudinal theory of pleasure might say that to have pleasure at t is to have some feeling F at t and to want to be feeling F at t . The question is how someone who holds this kind of view might account for the explanandum, which, again, is this:

Explanandum: We sometimes become aware, at least partly through phenomenology, that we are in a state of pleasure or pain.

It is not exactly clear what it would mean to become aware that one is in a state of pleasure or pain in the attitudinal sense. I do not think it needs to involve any explicit thought along the lines of "I want to be feeling what I am feeling now". But neither is it enough for the relevant attitude to merely be present. The subject must come into some form of

mental contact with it, such that a state in which the attitude is present is subjectively distinguishable from a state in which it is absent.

We know that we can come into contact with some of our mental states through phenomenology. But we should consider that there might be another way in the case of attitudes. Here an attitudinal theorist might point to a passage from Alston (1967). Writing in support of an attitudinal theory of pleasure, and anticipating an argument like the one I am giving now, Alston writes:

[T]here are many things to which an individual has privileged access that cannot be regarded as immediately felt qualities, such as intentions, attitudes, and beliefs. If I intend to quit my job tomorrow, I know that I have this intention without having to do any investigation to find out; I know just by virtue of having the intention; I know this as immediately as I know that I am now aware of a reddish patch . . . Yet an intention is neither a felt quality nor a complex of felt qualities. Hence the epistemological status of pleasure is not a conclusive reason for construing it as a quality of experience . . . among the nonsensory quality items to which a person has privileged access are his likes, preferences, and wants. It seems reasonable to suppose that a person's knowledge that he would choose to have an experience just like his present one on the basis of its felt quality can be just as immediate as his knowledge that he is aware of a red patch. (p. 345)

Taking inspiration from this, an attitudinal theorist could propose the following: when we become aware of our hedonic states, we come into contact with the phenomenal aspect of the state through its phenomenology, and with the requisite attitude through a kind of direct access unmediated by phenomenology.¹⁸ Call this the “direct access account”. On

¹⁸ I am not certain that Alston means to propose that we can have privileged access to our attitudes, intentions, or beliefs in a manner entirely unmediated by phenomenology. I think this is a reasonable interpretation of the passage, but admittedly, Alston does not say it outright. This is why, in the main text, I merely say that an attitudinal theorist could propose this view, and point to Alston as inspiration.

That said, here is why I think it is a reasonable interpretation of the passage. Alston contrasts privileged access to immediately felt qualities with privileged access to “intentions, attitudes, and beliefs”. He follows with an example of the latter: knowing that he intends to quit

this view, phenomenology is indeed doing part of the work in bringing about awareness of a hedonic state, so the explanandum is accounted for.

In other respects, however, the direct access account is not tenable. In the above quotation, Alston speaks of knowledge rather than awareness, and it may be possible to have the former without the latter. But I deny that we can be aware of mental states that do not affect us phenomenologically.

I deny this on the following grounds. Recall that to be aware of X is to consciously know that X. So, to be aware that you have some attitude is not just to know that you have that attitude, but to consciously know that you have that attitude. And you cannot consciously know that X without having the fact that X consciously before your mind, i.e. before your mind in a way that makes a difference to what it is like to be you at that moment. This is just to say that conscious knowledge necessarily involves some phenomenology, namely the feeling of having the known proposition consciously before your mind. So, by definition, you can only become aware of something if it makes a difference in your phenomenological field (directly or indirectly). But the direct access

his job tomorrow. And he contrasts this with knowledge that he is now aware of a reddish patch. From this, I infer that being aware of a reddish patch is meant to be an example of an immediately felt quality. Later, in the final sentence quoted above, Alston says that “[i]t seems reasonable to suppose that a person's knowledge that he would choose to have an experience just like his present one on the basis of its felt quality can be *just as immediate* as his knowledge that he is aware of a red patch” (emphasis mine). This is to say that knowledge of one of your attitudes can be just as immediate as knowledge of one of your immediately felt qualities. But if knowledge of one of your attitudes was necessarily mediated by phenomenology—that is, necessarily mediated by an immediately felt quality—then knowledge of an attitude could not be just as immediate as knowledge of an immediate felt quality. So Alston does appear to be proposing that we can have a kind of privileged access to our attitudes that is unmediated by phenomenology.

account implies that we can become aware of the onset of a pleasure or pain while undergoing no change whatsoever in what we are feeling. This is not coherent.

So, for an attitudinal theory to account for the explanandum, it must posit that the relevant attitude has a phenomenology of its own (in it at least some cases). This would allow phenomenology to put us into contact with both of the mental states involved in the relevant cases of pleasure and pain: feelings, and attitudes held toward those feeling.

Lin (2019) has proposed a theory of pleasure along these lines, which he calls the “hybrid theory”:

Lin’s Hybrid Theory: There is a kind of favorable attitude, A, that is partly constituted by a certain phenomenology, P. An attitudinal pleasure is an experience consisting, at least in part, of your tokening A toward a state of affairs. A sensory pleasure is an attitudinal pleasure whose object is an obtaining state of affairs consisting of your presently experiencing a particular sensation. (p. 519)¹⁹

On this view, every instance of pleasure involves a favorable attitude, and it always feels a certain way for the subject of the pleasure to have that attitude. Lin is noncommittal about what the relevant favorable attitude might be, but for illustration, let’s say it is the attitude of desire. In this case, the hybrid theory would assert that part of what it is to desire something is to experience a certain feeling. Consequently, on this view, in every

¹⁹ Lin puts this view forth as “a way of reconciling or integrating attitudinal and phenomenological theories of pleasure”, such that “the main claims of both types of theory are true: pleasures are pleasures in virtue of how they feel, and pleasures are pleasures in virtue of how they are related to the favorable attitudes of the subjects who experience them” (p. 518). This integration of phenomenological and attitudinal aspects may be the source of the term “hybrid”.

Lin is clear that the attitude A, which is partly constituted by the phenomenology P, is also partly constituted by some non-phenomenal aspects, saying that “to token this attitude toward a state of affairs is, *among other things*, to feel a certain way about that state of affairs” (p. 519, emphasis mine). He does not say why A could not be wholly constituted by P, that is, why the relevant attitude could not just be a kind of feeling. I briefly discuss this proposal below.

case of sensory pleasure there are at least two distinct phenomenologies: the phenomenology that is partially constitutive of desire, and the sensation at which the desire is directed.

I believe this approach to hedonic states is mistaken. Since the distinction between attitudinal and sensory hedonic states is not relevant in this particular context, I will explain why I reject hybrid theories of pleasure and pain with reference to these two statements:

Hybrid Theory of Pleasure: There is a kind of favorable attitude, FA, that is partly constituted by a certain phenomenology, P. A pleasure is an experience consisting, at least in part, of your tokening FA toward something you are currently feeling.

Hybrid Theory of Pain: There is a kind of disfavorable attitude, DA, that is partly constituted by a certain phenomenology, P*. A pain is an experience consisting, at least in part, of your tokening DA toward something you are currently feeling.

The problem is not that hybrid theories cannot account for the explanandum. On the contrary, these theories do account for the explanandum (with respect to pleasure and pain, respectively). On these views, the favorable and disfavorable attitudes (FA and DA), as well as the feelings at which they are directed, are at least partially constituted by their respective phenomenologies. So, phenomenal introspection should be enough to put us in touch with all of them in a manner sufficient for conscious awareness of the resulting pleasure or pain.

The weaknesses of hybrid theories lie elsewhere. To see how, consider the following pair of challenges.

First, note that because the pro-attitude phenomenology P is only partly constitutive of the pro-attitude FA, and the disfavorable attitude phenomenology P* is

only partly constitutive of the disfavorable attitude DA, it must be logically possible for P to be present in the absence of FA and P* to be present in the absence of DA. In other words, it must be logically possible for you to feel as though you have the relevant favorable attitude toward what you are feeling without it actually being the case that you have this attitude, and so too for the relevant disfavorable attitude.

So, a hybrid theory of pleasure must allow, as a matter of logical if not physical possibility, that when you take yourself to be feeling pleasure from your self-administered scalp massage (or other pleasure-inducing activity), you might not actually be having any pleasure at all—even though your experience in that moment is phenomenologically indistinguishable from pleasure, right down to the feeling of having a favorable attitude toward the massaging sensation (or other seemingly pleasurable sensation). Similarly, a hybrid theory of pain must allow, for example, that when you take yourself to be feeling pain from your self-administered pinch, you might not actually be in pain at all—even though your experience in that moment is phenomenologically indistinguishable from pain, right down to the feeling of having a disfavorable attitude toward the pinching sensation.

By the same token, the hybrid theory of pleasure must allow, as a matter of logical if not physical possibility, that you could have a fantastic orgasm, feel everything you would be feeling if you were in a state of tremendous physical pleasure—including the feeling of having a strongly favorable attitude toward your current bodily sensations—and yet not be in a state of pleasure at all. Similarly, the hybrid theory of pain must allow that, as a matter of logical if not physical possibility, you could shatter one of your legs in

a car accident, feel everything you would be feeling if you were in terrible agony—including the feeling of having a strongly disfavorable attitude toward the intense sensations emanating from your injured limb—and yet not be in any pain at all. In both cases, this would be possible because FA and DA are each only partly constituted by their respective phenomenologies (P and P*); it must therefore be at least logically possible to experience these phenomenologies in the absence of whatever else is required for tokening the corresponding attitude. I submit that these implications are implausible enough that we should reject the hybrid approach. (If you do not agree, then this challenge will not convince you.)

The second challenge begins with the observation that, in order to avoid an infinite regress, P (the phenomenology that is partly constitutive of FA) must not be pleasurable in itself, and P* (the phenomenology that is partly constitutive of DA) must not be painful in itself. On the hybrid view of pleasure, P could only be a pleasure if it were an experience consisting (at least in part) of the subject tokening FA toward something he is feeling, where that token of FA is itself partly constituted by a pleasant phenomenology, and so on *ad infinitum*. So a hybrid theory of pleasure must hold that P, by itself, is hedonically neutral. Similarly, on the hybrid view of pain, P* could only be painful if it were an experience consisting (at least in part) of the subject tokening DA toward something he is feeling, where that token of DA is itself partly constituted by a painful phenomenology, and so on *ad infinitum*. So a hybrid theory of pain must hold that P*, by itself, is hedonically neutral. But this has an implausible implication when it comes to our explanandum.

To see what I mean, recall our earlier experiment with arm-pinching. When you performed this experiment, you experienced a transition between awareness of a non-painful feeling (the pinching feeling before it became painful) and awareness of a painful feeling (the pinching feeling after it became painful). The hybrid theory of pain suggests that you were able to do this, via phenomenology, because you were able to come into the right sort of conscious contact with both (a) a pinching sensation in your arm and (b) P*, the phenomenology partly constitutive of the disfavorable attitude directed at the pinching sensation. Both of these parts, sensation and attitude, were necessary for pain to occur; on the hybrid theory (and indeed any attitudinal theory) of pain, you could have felt the exact same sensation without being in pain at all. Pain began with the arrival of the disfavorable attitude, and your awareness of your pain became possible when you began feeling P*. And as we've seen, P* must be hedonically neutral in itself. So the hybrid theory implies that what distinguished pain from non-pain for you, introspectively speaking, was the onset of a hedonically neutral feeling.

I submit that this does not track with what we observe, introspectively, when we perform the arm-pinching experiment. We do not become aware of pain by noticing that an extra, hedonically neutral phenomenology has appeared alongside the pinch. This just does not describe what it is like for a sensation to start hurting. (If you disagree with this judgment, then this line of argument will not persuade you.)

If you had success with giving yourself a pleasant scalp massage, or were otherwise able to induce pleasure in yourself in a manner roughly equivalent gradually inducing pain via arm-pinch, then we can run the same sort of argument for pleasure.

When you performed this experiment—let’s say you did give yourself a scalp massage—you experienced a transition between awareness of a non-painful feeling (the massaging feeling in the brief moment before it became pleasurable) and awareness of a pleasure (the massaging feeling after it became pleasurable). The hybrid theory of pleasure suggests that you were able to do this, because you were able to come into the right sort of conscious contact with both (a) a massaging sensation on your scalp and (b) P, the phenomenology partly constitutive of the favorable attitude directed at the massaging sensation. Both of these parts, sensation and attitude, were necessary for pleasure to occur; on the hybrid theory (and indeed any attitudinal theory) of pleasure, you could have felt the exact same sensation without experiencing any pleasure at all. Pleasure began with the arrival of the favorable attitude, and your awareness of your pleasure became possible when you began feeling P. And as we’ve seen, P must be hedonically neutral in itself. So the hybrid theory implies that what distinguished pleasure from non-pleasure for you, introspectively speaking, was the onset of a hedonically neutral feeling.

I submit that this does not track with what we observe, introspectively, when we perform the scalp-massaging (or other pleasure-inducing) experiment. We do not become aware of pleasure by noticing that an extra, hedonically neutral phenomenology has appeared alongside the massaging sensation. This just does not describe what it is like for a sensation to start feeling pleasant. I submit that these implications are implausible enough that we should reject the hybrid approach. (And again, if you disagree with this judgment, then this line of argument will not persuade you.)

These challenges exploit basic features of hybrid theories of hedonic states: their claim that each hedonic state requires both a phenomenology and an attitude directed at that phenomenology, and their claim that the attitude is itself partly constituted by a distinct phenomenology of its own. Every hybrid theory of hedonic states will make claims of these kinds, as they form the attitudinal and phenomenological components of the titular hybrid, respectively. So, versions of my two challenges will apply, *mutatis mutandis*, to any hybrid theory of hedonic states, not just ones I have used for illustration here.

They will also apply to theories on which the relevant attitude does not necessarily have any distinct phenomenology, but is accompanied by a tell-tale phenomenology often enough that the latter is a reliable indicator of the former. Call these sorts of attitudinal theories of hedonic states “coincidence theories”:

Coincidence Theory of Pleasure: There is a favorable attitude, FA, that is reliably accompanied by a certain phenomenology, P. A pleasure is an experience consisting, at least in part, of your tokening FA toward something you are currently feeling; this will usually, though not necessarily, coincide with feeling P.

Coincidence Theory of Pain: There is a disfavorable attitude, DA, that is reliably accompanied by a certain phenomenology, P*. A pain is an experience consisting, at least in part, of your tokening DA toward something you are currently feeling; this will usually, though not necessarily, coincide with feeling P*.

Like the hybrid theories of pleasure and pain considered earlier, coincidence theories must allow that you could feel P and P* in the absence of FA and DA. This gives us the same implausible implications as before: all the feelings of orgasm, including pro-attitude phenomenology P, without pleasure, and all the feelings of a shattered leg, including disfavorable attitude phenomenology P*, without pain. Moreover, coincidence theories

must hold, on pain of infinite regress, that FA is not necessarily pleasurable and DA is not necessarily painful. This commits coincidence theories to the notion that, in our arm-pinching and scalp-massaging experiments, we become aware of the onset of hedonic state upon the arrival of an additional, hedonically neutral feeling. Thus we should reject coincidence theories for much the same reasons that we should reject hybrid theories.

There is one last sort of theory left to consider: views on which the relevant attitude is not just partly, but wholly constituted by some phenomenology. Call these “fusion theories”:

Fusion Theory of Pleasure: There is a kind of favorable attitude, FA, that is wholly constituted by a certain phenomenology, P. A pleasure is an experience consisting, at least in part, of your tokening FA toward something you are currently feeling.

Fusion Theory of Pain: There is a kind of disfavorable attitude, DA, that is wholly constituted by a certain phenomenology, P*. A pain is an experience consisting, at least in part, of your tokening DA toward something you are currently feeling.

Notice that if one accepts a fusion theory of hedonic states, then one accepts the phenomenological thesis: that all pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures (with the same being true, *mutatis mutandis*, of pains). So fusion theories have no problem accounting for the explanandum, and they escape the objections I have pressed against hybrid and coincidence theories. But this is because they are consistent with my thesis. And for that very reason, they invite the same objections against phenomenological theories that I will be addressing below. Moreover, they face the additional problem of how to make sense of the idea that a wholly phenomenological state could be tokened toward states of affairs. (Can phenomenologies have propositional contents?) So while I

will allow that fusion theories are attitudinal theories of a sort, and I will not argue against them here, I do regard them as weakly motivated.

I conclude that there is no plausible way to account for the explanandum that is consistent with an attitudinal theory of hedonic states, except perhaps by adopting a view on which the relevant attitudes just are phenomenologies. But that's not all. We can extend my challenges to hybrid and coincidence theories to any other sort of non-phenomenological view of hedonic states. Any theory on which hedonic states are partly defined by something non-phenomenal will face the same fundamental problems.

Call this non-phenomenal prerequisite for the occurrence of a hedonic state "NP". (To illustrate: on a desire-based theory of pleasure, where desire is taken to be at least partly constituted by a behavioral disposition, NP would be that disposition.) In order to account for our explanandum, NP will have to have, on at least some occasions, a distinct, introspectible phenomenology; otherwise, it will not be the sort of thing we could become aware of. It will follow that it is logically possible for this phenomenology to be present in the absence of NP, and therefore also possible to have something that feels exactly like an extremely intense hedonic state—right down to the phenomenology associated with NP—without being one. Moreover, the distinct phenomenology of NP will have to be hedonically neutral in itself, on pain of infinite regress (just as we saw with P and P* above). It will follow that our pleasures and pains become phenomenally introspectible only with the presence of an extra, hedonically neutral phenomenology. These results will be just as implausible no matter what NP is taken to be.

2.2.3 Phenomenological Alternatives

If what I have said so far is correct, then we should reject non-phenomenological accounts of hedonic states. It follows that we should think of pleasures and pains as feelings that do not gain their status as pleasures or pains from their relations to non-phenomenal properties (such as attitudes). But it does not follow that we should embrace the phenomenological thesis. There remains an alternative phenomenological approach, on which all pleasures and pains are feelings, but neither pleasures nor pains all feel alike in some single defining respect.

This alternative needs some account of what unifies pleasure and pain as categories, if it is neither a relation to the subject's attitudes (since this is a phenomenological theory) nor a single phenomenal property (since this approach rejects the phenomenological thesis of BPH).²⁰ For this, I can see only one remaining option: posit that pleasures and pains are each unified as categories by similarities among their various phenomenal properties. For example, to borrow Wittgenstein's (1953: p. 66) famous phrases, pleasures and pains could be unified by "family resemblances" defined by "a complicated network of similarities overlapping and criss-crossing". In this case, the resemblances would be phenomenological. The idea would be that no two pleasures or two pains need share a single defining phenomenal property, but would share a "family resemblance" in the way they feel.

²⁰ See Goldstein (1985) for criticism of purportedly pluralistic theories of hedonic states that fail to provide a coherent account of what unifies pleasure and pain as categories.

Call this view, on which pleasures and pains are unified as categories by phenomenal similarities rather than single phenomenal properties, the “disunified phenomenological thesis” (DPT). Assuming the DPT is true, we could account for the explanandum in the following way. We could propose that, as we grow up, we notice phenomenal resemblances among various feelings, and at least partly on that basis, form our concepts of pleasure and pain. Subsequently, we become aware that certain of our occurrent feelings are pleasures or pains by introspecting and detecting the phenomenal properties that mark them as such. This could happen without the categories of pleasure or pain being defined by any single phenomenal property that all category members share.

I reject the DPT on the grounds that it is more complicated than the phenomenological thesis without having any distinct advantages. I will attempt to show that it is more complicated, in an epistemically relevant sense, at §4.2.2. As for distinct advantages, I can think of two facts that a proponent might claim the DPT can better explain than the phenomenological thesis: the fact that there are cases in which people judge that two pleasures do not feel alike in any way (and so too for pains), and the fact that we are sometimes uncertain as to whether a given feeling of ours is a pleasure, a pain, or neither.²¹

I will address the question of whether, and how, we can account for judgments of phenomenal disunity about hedonic states without giving up the phenomenological

²¹ Thanks to David Wong (personal communication) for calling this second point to my attention.

thesis in §2.3. As for cases in which we are uncertain whether a given feeling is a pleasure, a pain, or neither, we can account for these in a manner consistent with the phenomenological thesis by proposing that these are cases of phenomenal complexity, in which the subject's phenomenological field contains (a) both pleasure and pain in roughly equal measure, (b) mild pleasure or pain alongside other phenomenal properties of roughly equal or greater intensity, such that the hedonic states are difficult to introspectively discern, (c) phenomenal properties that are typically accompanied by pleasure or pain, but are not in this case, or (d) some perplexing combination of the above. This is speculative, of course, but it is enough to show that the phenomenological thesis is compatible with there being cases in which we don't know quite how to categorize what we are feeling.

This concludes my positive argument for the phenomenological thesis. I will now defend this thesis from some counterarguments.

2.3 General Arguments Against the Phenomenological Thesis

Once again, the phenomenological thesis is this:

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

In short, the thesis claims that both pleasures and pains are phenomenally unified.

Many philosophers deny this, and all that I have found deny it in a similar way: by claiming that introspection tells against it. To impress upon the reader how common this objection is, and how confidently it is typically stated, I will quote six examples at some length.

First, here is Sidgwick (1981: p. 127) on pleasure:

[W]hen I reflect on the notion of pleasure—using the term in the comprehensive sense I have adopted, to include the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments—the only common quality that I can find in the feelings so designated seems to be that relation to desire and volition expressed by the general term “desirable”. I propose therefore to define Pleasure—when we are considering its “strict value” for the purposes of quantitative comparison—as a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable, or—in cases of comparison—preferable.

Here is Alston (1967: p. 344), also reflecting broadly on pleasure:

Nevertheless, on further probing, the thesis that pleasure is a quality that can attach to any state of consciousness is not very plausible phenomenologically. When we reflect on a wide variety of cases of getting pleasure . . . we are unable to isolate a felt quality which they all share, in the way in which we can easily isolate a quality of redness which a number of different visual sensations share On the contrary, enjoying playing tennis feels very different from getting satisfaction out of seeing an enemy in distress, and both feel very different from the sense of well-being one has when, in good health, one arises carefree from a good night's sleep.

And Sobel (2002: p. 241):

The first objection to quantitative hedonism is that there is no single sensation that is common to all our different experiences of intrinsic value. The pleasures of walking barefoot through the grass arm in arm with your love have so little phenomenologically in common with the pleasures of winning a tense tennis match or eating a good burger or working through a challenging philosophical problem that we do not understand the instruction to maximize the sensation that these different activities share.

. . . Many have introspected in vain searching for this alleged experiential commonality among the full array of human pleasures.

Feldman (1997: pp. 83-84) comments on the narrower topic of sensory pleasure:

One thing to notice about sensory pleasure is its apparent heterogeneity. The man on the beach enjoys some pleasurable smells as well as some pleasurable feelings of warmth. Each of these sensations is pleasant, pleasurable, 'pleasure-giving'. Some would find nothing odd in saying that each of these sensations 'is a pleasure'.

Nevertheless, from the strictly phenomenological perspective, they seem to have very little in common. One is an olfactory sensation - it is the smell of fresh, salty air. The other is an all-over bodily feeling of warmth. Aside from the fact that they are experienced simultaneously and by the same person in the example, they seem to be utterly unlike.

In order to see the heterogeneity of sensory pleasures even more clearly, consider the pleasurable sensations you get when you eat delicious, salty peanuts and drink sparkling, cold beer. The taste of the peanuts is a pleasure. The taste of the beer is a pleasure. Yet, unless your taste sensations are profoundly unlike mine, the taste of the peanuts has little in common with the taste of the beer. Indeed, it is reasonable to suppose that we who love to eat peanuts while we drink beer love this, at least in part, because of the remarkable contrast between the two leading sensations involved. (The phenomenological contrast between the taste of cold beer and the feelings of warmth enjoyed while sunbathing is even more striking - yet each of these may be a sensory pleasure.)

On pain, meanwhile, here is Korsgaard (1992: pp. 147-148):

[I]f the painfulness of pain rested in the character of sensations . . . our belief that physical pain has something in common with grief, rage and disappointment would be inexplicable. For that matter, what physical pains have in common with each other would be inexplicable, for the sensations are of many different kinds. What do nausea, migraine, menstrual cramps, pinpricks and pinches have in common, that makes us call them all pains? (Don't say they're all horrible; that's just repeating yourself.)²²

Finally, Parfit (1984: p. 493) denies phenomenal unity to pleasure and pain both (emphasis in original):

Narrow Hedonists assume, falsely, that pleasures and pains are two distinctive kinds of experience. Compare the experience of satisfying an intense thirst or lust, listening to music, solving an intellectual problem, reading a tragedy, and

²² In the Feldman, Alston, Sobel, and Parfit passages, the invitation to reflect on one's pleasures is more or less explicit. Sidgwick reports his own reflections, but it's reasonable to infer an invitation to the reader to reflect similarly. On the surface, Korsgaard might appear to be doing something else, namely asserting that the phenomenal unity of pain would render some fact inexplicable. But the (alleged) fact here just is that pains are phenomenally heterogeneous. Korsgaard's only support for this claim comes from listing examples of pains. I interpret this too as an appeal to introspection, with Korsgaard implicitly inviting the reader to reflect on their experiences of grief, rage, nausea, headaches, and so on.

knowing that one's child is happy. These various experiences do not contain any distinctive common quality.

What pains and pleasures have in common are their relations to our desires. On the use of 'pain' which has rational and moral significance, all pains are when experienced unwanted, and a pain is worse or greater the more it is unwanted. Similarly, all pleasures are when experienced wanted, and they are better or greater the more they are wanted.

The argumentation in these passages is not very explicit. But each of them at least hints at a straightforward argument with the following structure:

P1. If not all [pleasures/pains] share a distinctive phenomenal quality, then [pleasures/pains] are not phenomenally unified.

P2. Not all [pleasures/pains] share a distinctive phenomenal quality.

C. [Pleasures/pains] are not phenomenally unified.

In each case, P2 is supported by appeal to phenomenal introspection. (P1 is true by definition.) Each philosopher invites the reader to reflect on various pleasures or pains and attempt to discern in them a common phenomenal quality. Presumably, the idea is that the reader will find no such quality, which is evidence of there being no such quality.

This strategy of supporting P2 by appeal to phenomenal introspection rests on two assumptions. The first assumption is that the reader will make the expected introspective judgments. How often this is true I cannot say, so I will grant for the sake of argument that virtually everyone does so (though I myself do not).

The second assumption is that we are justified in taking phenomenal introspection—which, in this context, I take to encompass both introspection of present feelings and recollection of past feelings—to be a sufficiently reliable way of determining whether different hedonic states share a distinctive phenomenal quality. By “sufficiently

reliable”, I mean reliable enough to make our beliefs about which hedonic states share distinctive phenomenal qualities at least *prima facie* justified. In what follows I will use the term “reliable” as shorthand for reliability in this sense.

In other words, for this way of refuting the phenomenological thesis to work, we must be justified in believing that we have what I will call “qualitative insight”:

Qualitative Insight: A subject has qualitative insight—is qualitatively insightful—iff they can use phenomenal introspection to reliably determine whether or not various hedonic states share a distinctive phenomenal quality.

If we are qualitatively insightful, and we cannot detect any distinctive phenomenal quality among pleasures or among pains, then the phenomenological thesis is false. It is important, then, to assess what reasons we have to accept the claim that we have qualitative insight. But, despite this claim being neither self-evident nor uncontroversial, I have yet to find an argument for it, in the work of the above-quoted philosophers or any others. In the absence of such an argument, the sort of strategy employed by Sidgwick et al above cannot refute the phenomenological thesis.

In the interest of charity, I will consider two rationales for accepting the claim of qualitative insight—the sorts of rationales a philosopher might regard as too obvious to be worth mentioning—and try to show that each is mistaken. The rationales are these: that only by assuming qualitative insight can we account for our introspective judgments about (the lack of) phenomenal unity among hedonic states; and that the claim of qualitative insight is the default position about our ability to introspect pleasures and pains, such that the burden is on its opponents to give reasons against it.

First, the matter of accounting for our introspective judgments. Again, I have assumed for the sake of argument that these judgments cut against the phenomenological thesis. On this assumption, they would include, for example, judgments to the effect that a pleasurable taste of peanuts and a pleasurable taste of beer share no single phenomenal quality that marks them both as pleasures, and that the feeling of nausea and the feeling of grief share no single phenomenal quality that marks them both as pains. One rationale for accepting the claim of qualitative insight might be that without that claim, we can't explain why we would make these judgments. If that were the case, it would give us compelling reason to believe that we are qualitatively insightful, which would then underwrite the crucial premise (P2) of introspective arguments against the phenomenological thesis.

But there are other ways, consistent with the phenomenological thesis, of explaining why we would make these judgments. I will provide two such accounts: one for the separate experience view of hedonic states, and one for the hedonic tone view (each introduced in §2.1 above).

The separate experience view of hedonic states proposes that pleasure and pain are each a distinctive kind of feeling—the feeling of pleasure itself and the feeling of pain itself—which could, in principle, be felt in isolation from anything else. This would mean that each occurrent hedonic state is just an instance of one of these feelings. On this view, it is not so much that pleasure and pain have distinct phenomenal qualities. Rather, they simply are phenomenal states, defined by their distinct phenomenal qualities.

Now, if it were common to actually feel a hedonic state in phenomenal isolation—which the separate experience theory allows is logically possible—then it might well be easy to introspect the distinct phenomenal qualities of pleasure and pain. So this view should stipulate that, as a matter of contingent physical fact, pleasures and pains are virtually always felt alongside other qualia. So, for example, when you feel pleasure from tasting something delicious, the pleasure overlaps in consciousness with the various flavor qualia; and when you feel pain from stubbing your toe, the pain overlaps in consciousness with the stinging and throbbing sensations that seem to be emanating from your foot. This forecloses one way in which the separate experience view might be thought to predict that we would judge pleasures and/or pains to be phenomenally unified.

This stipulation also allows the separate experience view to account for the impression that pleasures and pains are phenomenally heterogeneous. The theory denies that pleasures and pains are heterogeneous themselves, but can affirm the heterogeneity of the other qualia with which hedonic states overlap in consciousness. A proponent of the separate experience view need not deny that taste of peanuts is different from the taste of beer, or that the feeling of nausea is different from the feeling of grief, and indeed, he can point to the heterogeneity of non-hedonic qualia to explain why we would make the mistake of judging that there are pleasures and pains that do not feel at all alike.

A proponent of the separate experience view should also propose that, as Bramble (2013: p. 210) puts it, hedonic states “occupy” or “permeate” the other qualia with which they overlap, in such a way that introspection cannot cleanly separate the phenomenal

contributions of these states from those of the other qualia being experienced. If introspection could suss out pleasure and pain in this way, then we would be qualitatively insightful, and our introspective judgments would give decisive evidence against the separate experience theory. So, on pain of self-defeat, the theory must stipulate that pleasure and pain intertwine themselves with other qualia so as to evade this kind of introspective detection (without, of course, rendering us unable to distinguish pleasures from non-pleasures and pains from non-pains).

This claim also allows the separate experience theorist to account for the impression that a single sort of experience—that of a flavor, for example—can, for a single person, be pleasant on one occasion and unpleasant on another. If pleasure and pain permeate the other qualia with which they overlap, then these are cases in which some qualia are accompanied (and phenomenologically permeated) by the feeling of pleasure itself on one occasion and the feeling of pain itself on the other.

Thus we can assume the separate experience view of hedonic states, account for judgments of phenomenal disunity among pleasures and among pains, and not run afoul of common sense. This is done by proposing that the feelings of pleasure and pain are, in practice, always felt at the same time as other qualia, which they permeate in a way that stymies qualitative insight.

We can account for our introspective judgments on the hedonic tone view of pleasure and pain as well. This view also affirms the phenomenological thesis, but denies that pleasure and pain can be felt in isolation from other qualia. It proposes instead that

hedonic states occur when (and only when) some among a subject's occurrent qualia take on a pleasurable or painful tone.

The hedonic tone view allows that peanuts can taste very different from beer, even when both tastes are pleasures; nausea can feel very different from grief, even as both states are pains; and so on. It also allows that some qualia can, for a single subject, have a pleasurable tone on one occasion and a painful one on another. This is compatible with these states being unified, as pleasures and pains respectively, by these hedonic tones. And it may be the case that, when we use introspection to compare pairs of pleasures or pairs of pains, the hedonic tone that the members of each pair share eludes us—especially when we compare states that, beyond their common hedonic tone, are otherwise qualitatively disparate. Thus the hedonic tone account can also square phenomenal unity with introspective judgments to the contrary.

Of course, there are many further questions we could ask about these views. But the point is just to show that without giving up the phenomenological thesis, we can, in more than one way, explain why we would make introspective judgments like those that Sidgwick et al appeal to. This shows that the first rationale for believing that we are qualitatively insightful fails.

The second rationale for the claim of qualitative insight presents this claim as the default position about our ability to introspect pleasures and pains, such that the burden is on its opponents to give reasons against it. One might propose this second rationale on the grounds that we must grant *prima facie* epistemic credibility to our introspective capacities in order to ward off skepticism about introspective knowledge. Perhaps this is

true. But I believe that my positive argument for the phenomenological thesis does give us reason to reject the claim of qualitative insight. If I'm right about this, then it would not matter if belief in qualitative insight were the default position, because I have already shouldered the burden of giving reasons against it. And there is nothing skeptical about declining to believe that some capacity is reliable when you have positive reason to believe that it isn't.

I conclude that introspective arguments against the phenomenological thesis like those given (or suggested) by Sidgwick, Alston, Sobel, Feldman, Korsgaard, and Parfit do not give us any reason to reject the phenomenological thesis. To refute this thesis directly, opponents need arguments that don't rely on the claim that we are qualitatively insightful. I have yet to encounter any such arguments.

2.4 A General Argument Against the Calculation Thesis

Now on to a general argument against the calculation thesis:

Calculation Thesis: All pleasures have magnitudes that are quantifiable (in principle, using cardinal numbers) and fully commensurable; the same is true of pains.

The calculation thesis tells us that there is total quantitative continuity (hereafter simply "continuity") among the magnitudes of pleasures, and so too for pains. This is to say that for any given hedonic state, there is a quantifiable fact of the matter as to how large its magnitude is, and for any two pleasures or two pains, there is an exact numerical ratio between their magnitudes. (In other words, for any two token pleasures A and B, there is

a cardinal number n such that A contains exactly n times as much pleasure as B.²³) An argument against the calculation thesis, then, must seek to show that there is quantitative discontinuity (hereafter simply “discontinuity”) among hedonic states, i.e. that there are cases in which two pleasures or two pains lack a precise numerical ratio between their magnitudes.

Having already addressed arguments against the phenomenological thesis, we need not consider arguments that infer discontinuity from phenomenal disunity. We must also avoid mistaking arguments for the practical impossibility of calculating the numerical ratios between hedonic magnitudes for arguments against the calculation thesis, which claims only that such calculations are logically possible. And given that this chapter concerns descriptive objections to BPH, it is not the place to discuss arguments for discontinuity that depend on evaluative premises (we will see two such arguments in chapter 3).

I have yet to find an argument in the literature that meets these conditions. Feldman (2004: pp. 45-49) points to a passage from Brentano (2009: pp. 30-31) as possibly giving an argument of this kind, but I believe he is mistaken.²⁴ Here is the passage (emphasis mine):

There are some who hold, in opposition to what experience makes evident to us, that pleasure is the only thing good in itself, that pleasure is the good. If this view were true, then, as Bentham urged, it would have the following advantage: since all goods would be homogeneous, we would be able to compare them quantitatively and thus determine their relative values...

²³ I owe this formulation to Feldman (2004: p. 41).

²⁴ In fairness, Feldman (2006: p. 47) says that he is “not entirely sure” about this interpretation. One of Feldman’s doctoral students, Klocksiem (2009: p. 134), interprets Brentano as rejecting the measurability of hedonic states, not their in-principle quantifiability or commensurability

But only a moment's reflection is needed to shatter such illusory hopes. *Is it really possible to find out whether one pleasure is twice as great as another? Gauss, who certainly knew something about measurement, has denied that this is possible.* A foot is divisible into twelve inches; but an intense joy is not divisible in the same sense into twelve less intensive joys. Consider how ridiculous it would be if someone said that the amount of pleasure he has in smoking a good cigar is such that, if it were multiplied by 127, or say by 1,077, it would be precisely equal to the amount of pleasure he has in listening to a symphony of Beethoven or in viewing one of Raphael's madonnas! This is enough, I think, to suggest the further difficulties involved in trying to compare the intensity of pleasure with that of pain.

The italicized section, which Feldman omits, makes clear that Brentano was pointing out our practical inability to measure hedonic magnitudes, not arguing for the non-existence of such magnitudes.

Still, given the rarity of general arguments against the calculation thesis, it worth considering how Brentano's remarks might be put to use in such an argument. Feldman suggests that they could be taken as an argument for the claim that pleasure magnitudes are incomparable in size, not only practically but in principle. Reconstructed as an argument against the calculation thesis, that argument would look something like this:

P1. If there is a pair of token pleasures, A and B, for which there is no number n such that A contains exactly n times as much pleasure as B, then the calculation thesis is false.

P2. There is a pair of token pleasures, A and B, for which there is no number n such that A contains exactly n times as much pleasure as B.

C. The calculation thesis is false.

This argument is similar to the arguments against the phenomenological thesis discussed earlier. It is so similar, in fact, that we can re-use the defensive strategy from §2.3 with only minor alterations.

As with the arguments discussed in §2.3, this argument is valid, P1 is true by definition, and the support for P2 comes from a comparison between pleasures, such as Brentano's comparison between the pleasure from smoking a cigar and the pleasure from listening to a symphony or looking at a painting. The idea would be that the absurdity of proposing any particular number for n here is evidence that there really is no such n . This could only seem absurd to us if we thought carefully about the comparison, which would require that we recall our own experiences with the sorts of pleasures being compared, or others that seem sufficiently similar. This is to say that, like the arguments in §2.3, this argument relies on appeal to phenomenal introspection.

As before, the strategy of supporting P2 by appeal to phenomenal introspection rests on two assumptions.

The first assumption is that the reader will make the expected introspective judgments. I accepted this assumption with respect to the arguments in the previous section, but I am reluctant to do the same here. I doubt that many people would have clear intuitions about the existence of precise numerical ratios between the intensities of any two token pleasures. What Brentano points out is that it would be absurd to declare a specific number, such as 127, as the exact ratio between the intensities of the pleasures gained from two experiences that are highly qualitatively distinct. But judging this to be absurd is not the same as judging that there is no such number identifiable even in principle. For the sake of argument, I will accept the assumption all the same—but only for comparisons between pleasant experiences that are highly qualitatively distinct (like the one Brentano suggests).

The second assumption is that we are justified in taking phenomenal introspection to be a reliable way of determining whether, for the pair of token experiences in question, there is a number n such that one contains exactly n times as much pleasure as the other. In other words, for this way of refuting the phenomenological thesis to work, we must be justified in believing that we have what I will call “quantitative insight”:

Quantitative Insight: A subject has quantitative insight—is quantitatively insightful—iff they can use phenomenal introspection to reliably determine whether, for any token pleasures A and B , there is a number n such that A contains exactly n times as much pleasure as B .

I know of no argument for this claim. And in the absence of such an argument, this sort of approach will not refute the calculation thesis.

Moreover, the comparisons most likely to elicit judgments in support of P2 would involve pleasant experiences that are in many respects phenomenologically unlike, as in the case of listening to a symphony and smoking a cigar. We can explain these judgments in a manner consistent with the calculation thesis being true, on both the separate experience and hedonic tone views of hedonic states, in the same way that we accounted for judgments that cut against the phenomenological thesis. On each view, we can explain the error by pointing out that the experiences being compared involve highly distinct qualia (e.g. the sound of violins and the taste of cigar smoke), and proposing that pleasure—conceived of as a separate experience or a hedonic tone—permeates these in consciousness in a way that stymies quantitative insight.

2.5 Separate Experience or Hedonic Tone?

Up to this point, I have not taken sides between the separate experience view and the hedonic tone view. Each of these theories embraces the phenomenological thesis, and each withstands the introspective arguments discussed in §2.3 above. But we might wonder whether one is preferable to the other. So, in this section, I will consider some representative objections to each view. In doing so, I will attempt to show that we should not favor the separate experience view over the hedonic tone view or *vice versa*. Instead, we should accept the phenomenological thesis while suspending judgment as to which sub-theory of this type is correct.

2.5.1 Idiosyncrasy Objections

The separate experience view holds that both pleasure and pain are feelings that can, in principle (though perhaps not in practice) be felt in isolation from other qualia. Some philosophers object to this view on the grounds that if it were true, hedonic states would be so unlike other feelings as to strain credulity. For example, here is Alston (1967: p. 342):

It would seem that any sensation, if it becomes sufficiently acute, will tend to monopolize consciousness and interfere with concentration on anything else. On the [separate experience] view under consideration, the more pleasure we get out of, say, playing the piano, the more intense the sensation of pleasure would become, the more our attention would be taken up with the sensation of pleasure, and the harder it would become to concentrate on the playing. But the reverse is the case. The more pleasure we get out of doing something, the easier it is to concentrate on it.

The objection here is that if pleasure is a separate kind of feeling (or sensation, as Alston puts it), it would stand alone among feelings in not becoming more distracting as it

becomes more intense. Stranger still, it would seem to be the one feeling that has the opposite effect.

In a similar vein, Sobel (2002: p. 242) says this:

[T]he sensations that we are familiar with can be either pleasant or unpleasant, agreeable or disagreeable, depending on the agent and the context. If we understand pleasure to be a particular flavor of sensation, then it would seem that the hedonist must claim that there is a flavor of sensation that will always benefit each one of us, regardless of our tastes. This seems radically at odds with our experience with flavors of sensations that we are confident are sensations.

Sobel objects that, if pleasure is a separate feeling (or sensation, as he also puts it), and prudential hedonism is true, then pleasure would stand apart as the only “flavor of sensation” that benefits everyone regardless of their tastes.

Alston and Sobel appear to share a suspicion of idiosyncrasy. The idea seems to be that, all else being equal, a theory should assign as few unique properties to pleasure as possible. Pleasure (and pain) must have some unique properties to be distinct from other mental states. On the separate experience view, these would be phenomenal properties. But both Alston and Sobel point out ways in which the separate experience view threatens to burden hedonic states with further idiosyncrasies.

I am willing to agree that if the separate experience view is correct, then pleasure is idiosyncratic in the way Alston suggests; that if both the separate experience view and prudential hedonism are true, then pleasure is idiosyncratic in the way Sobel suggests; and that a theory of hedonic states should assign as few unique properties to pleasure (and pain) as possible, all else being equal. The question is whether all else is indeed equal. At this point, we are considering whether to adopt the separate experience view or the

hedonic tone view; in this context, then, charges of idiosyncrasy against the separate experience view will only stick if the same (or equivalent) charges cannot be made, with equal justice, against the hedonic tone view.

But the same, or equivalent, idiosyncrasy objections could in fact be leveled against the hedonic tone view. The hedonic tone view claims that pleasure and pain cannot be felt separately from other qualia, but rather occur when some qualia take on a positive or negative hedonic tone, respectively. With respect to idiosyncrasy objections, this approach faces a dilemma. If there are no other phenomenal states that can occur only as qualities of other qualia—if there is no precedent for the kind of mental state that the hedonic tone view claims pleasure and pain to be—then clearly hedonic states would be idiosyncratic in this respect. But if there are other such phenomenal states, then Alston's and Sobel's objections might apply just as well to the hedonic tone view as the separate experience view.

Consider the question of precedent. As mentioned in §2.1 above, proponents of the hedonic tone view sometimes describe pleasure and pain as qualitative dimensions along which experiences vary. In this vein, Kagan (1992, p. 172) famously proposed (though did not fully endorse) an analogy with the loudness of sounds:

As an analogy, consider the loudness of auditory experiences—that is, sounds. It is obvious that loudness or volume is not a kind of sound. And it seems plausible to insist that loudness is not a single kind of component of auditory experiences. Rather, volume is a dimension along which sounds can vary. It is an aspect of sounds, with regard to which they can be ranked. Recognition of the qualitative differences between the sounds of a symphony, rain falling, and a bird chirping, does nothing at all to call into question our ability to identify a single dimension—volume—with regard to which these and other sounds can be ranked.

Similarly, then, pleasure might well be a distinct dimension of mental states, with regard to which they can be ranked as well. Recognition of the qualitative differences between the experiences of hiking, listening to music, and reading philosophy, need not call into question our ability to identify a single dimension—pleasure—along which they vary in magnitude.

In this way, Kagan locates a precedent for hedonic tone (conceived of in dimensional terms). Arpaly & Schroeder (2013: pp. 122-123), meanwhile, find precedent for a single continuous dimension of pleasure and pain in the qualitative dimension of warmth and cold.

But Bramble (2013: p. 209), a proponent of the separate experience view, objects that the loudness analogy is inapt. He points out that when an auditory experiences reaches the zero point on loudness dimension, it ceases to exist; but the same is not true of experiences on the pleasure dimension, which can continue even as their pleasantness is reduced to nothing. The same could perhaps be said against the analogy with warmth and cold: there may be no sensation which feels neither warm nor cold to any degree. (This is more dubious than the claim that there is no sound without some degree of loudness, but let's suppose that it's true.) So even if we agree that loudness, warmth, and cold provide precedent for the notion of a dimension along which experiences vary, hedonic tone will still be unique among qualitative dimensions in some way other than how the experiences on it feel. One might even claim that this idiosyncrasy is enough to mark the hedonic tone dimension as altogether *sui generis* among mental phenomena.

On the other hand, suppose we find Bramble's objection unconvincing, and regard the dimensions of loudness, warmth, and/or cold as adequate enough precedents for the dimensions of pleasure and pain. Then we are faced with the fact that versions of Alston's

and Sobel's objections apply just as well to the hedonic tone view as to the separate experience view. Sobel would surely object that, if prudential hedonism is true, then experiences along the pleasure dimension would stand alone as the only experiences which benefit us regardless of our tastes. And Alston would surely point out that the higher up on the loudness, warmth, or cold dimension an experience is, the more it tends to monopolize attention, and yet this is not so for experiences on the pleasure dimension. Hence the dilemma: the only way to save the hedonic tone view from Alston's and Sobel's objections is to make pleasure and pain so *sui generis* that they become vulnerable to an idiosyncrasy objection of another sort.

I conclude that even if we agree that a theory of hedonic states should assign as few unique properties to pleasure (and pain) as possible, all else being equal, this does not give us a reason to favor the separate experience view over the hedonic tone view or *vice versa*.

2.5.2 Epistemic Objections

Another objection to the separate experience view is that paints an implausible picture of when, and how, we know what is bringing us pleasure. For example, here is Kenny (2003: p. 90):

[I]f pleasure were a sensation its connection with what produced it would be a causal one. It would thus be only as a result of induction that we could say on any given occasion what we were enjoying. It would be possible to make exactly the same mistakes about what was giving one pleasure as it is possible to make about what has given one a stomachache. If, say, one had enjoyed listening to the first performance of a new overture, it would be a mere hypothesis that what one had enjoyed was listening to the overture and not, say, sitting in row G of the dress circle. This hypothesis would need to be verified in accordance with Mill's canons:

one should listen to the overture again, sitting in row F of the stalls, and introspect carefully to see if the same sensation occurred.

Feldman (1997: p. 342), who refers here to the separate experience view as the “Moorean” view, elaborates on Kenny’s argument:

Another difficulty is epistemic. Suppose I am drinking beer and eating peanuts more or less simultaneously. Suppose each of the tastes is pleasurable. Suppose, however, that the taste of the peanuts is more pleasurable than the taste of the beer. According to the Moorean view, here's what's happening. I am experiencing the taste of beer and the taste of peanuts. Each taste sensation is causing the feeling of pleasure itself. However, the taste of the peanuts is causing a more intense feeling of pleasure itself than is the taste of the beer.

If the Moorean view were true, I would face a certain slight difficulty when I tried to determine which taste is the greater pleasure. For, according to this view, I would be having four simultaneous relevant sensory experiences: the taste of the peanuts, the taste of the beer, and two feelings of pleasure itself, one more intensely than the other. While it would be reasonable for us to assume that the feelings of pleasure were being caused by the tastes, I might have to engage in some causal experimentation in order to determine which taste was causing which feeling of pleasure itself. Perhaps I would put aside the beer and munch on peanuts alone for a while. I could then check to see which feeling of pleasure persists. Then I might put aside the peanuts for a while, and take my beer straight. Once again, I could check to see which feeling of pleasure persists. With luck, I might be able to determine which taste sensation was the greater sensory pleasure.

The objection is that if pleasure were a separate experience—the feeling of pleasure itself—then in cases where we are having multiple experiences at once, some more pleasant than others, we would need to carry out some trial-and-error experimentation to determine which of these experiences are the more pleasant ones. But (the objection claims) this is not what happens. Alston (1967: p. 342) puts the point succinctly: “A person knows immediately which of the various things he is aware of at the moment he is taking pleasure in, and the sensation theory can give no account of this discrimination.”

I think this objection is mistaken. Kenny and Feldman are right to think that on the most plausible version of the separate experience view, the feeling of pleasure itself would be caused by the experiences we call pleasant or pleasurable. And they are right to point out that this implies a need for some trial-and-error reasoning to determine which of multiple overlapping experiences is the more pleasant one. The mistake is to suppose that this trial-and-error reasoning must take place consciously and deliberately, or involve anything as elaborate as listening to an overture again from a different row.

A proponent of the separate experience view could propose the following picture instead. First, the amount of pleasure that we feel from an experience, all else being equal, increases with our attention to that experience, and decreases when we direct our attention elsewhere. Second, when we have multiple overlapping conscious experiences, our attention tends to shift between them. For example, if we are listening to an overture from row G, our attention will tend to move from one aspect of the music to another, and perhaps in some moments will settle on the feeling of the seat underneath us, or the quality of our view of the orchestra. If we are eating peanuts and drinking beer, our attention will most likely shift from one flavor or texture to another, the peanuts being more salient when we bite into them, the beer more salient when it first hits the tongue. As this happens, we may detect some fluctuation in the amount of pleasure we feel. If there is a clear enough pattern in these fluctuations—say, a consistent burst of pleasure at the crunch of peanuts, or a consistent decrease in pleasure when we notice how poor the view is from row G—we will make an inference about which of our experiences is bringing us more pleasure and which less. (If two of our experiences are so blended in

consciousness that we cannot fix our attention more on one than the other, then presumably we would not make a judgment about which is more pleasurable.) All of this could happen without the sort of effortful experimentation that Kenny and Feldman seem to think would be required. Indeed, in normal cases it could happen so easily that we would be tempted to say, with Alston, that it happens immediately.

Admittedly, this is just more speculation on behalf of the separate experience view. But it is coherent, and neither Kenny nor Feldman gives us any reason to deny that this is how we typically know which of our overlapping experiences is more pleasant. I conclude that this sort of objection gives us no reason to reject the separate experience view.

Crisp (2006: pp. 104-105), meanwhile, poses an epistemic challenge to the hedonic tone view. The challenge concerns how we can distinguish between a component of experiences and an experiential dimension:

How is the distinction between components of experiences and dimensions of variation meant to work? Take the sound of a tinkling bell, and the sound of a honking horn. The components of each are, respectively, tinkling and honking. Volume, Kagan suggests, is not a 'kind' of sound. So a loud tinkling is the same sound as a soft tinkling, whereas a loud honk is a different sound from a loud tinkling.

It is questionable, however, whether this distinction captures anything of great metaphysical significance. We would indeed be inclined to say that the soft tinkling is the same sound as the loud tinkling. But that is because we usually focus on aspects of how things sound other than how loud they are. In fact loud sounds do form a kind. I might ask you to group sounds together according to their volume, and you would then categorize the loud tinkling with the loud honk, and the soft tinkling with the soft honk. As Kagan himself goes on to say, 'it seems. . . that there is a sense in which a specific volume is indeed an ingredient of a given sound'. Drawing distinctions between components, dimensions of variation, and ingredients of experiences does not seem a profitable direction in which to move.

If there is no genuine distinction between components of experience and dimensions of variation, and if this distinction is crucial for the hedonic tone view, then this is a serious challenge. As to the first question, I have no opinion. But as to the second, I think the answer is clear: it is not crucial for the hedonic view that there be a clean distinction between experiential dimensions and components of experiences. For a proponent of the hedonic tone view, what matters is that there be some precedent for the idea of a phenomenal property that varies continuously in magnitude and cannot be felt in isolation from other qualia. This is the point of the analogies with loudness, warmth, and cold. Whether there is a non-arbitrary way of dividing experiential dimensions from components is irrelevant (though perhaps continuous variation in magnitude could be the distinguishing criterion). Therefore, this objection does not give us any reason to favor the separate experience view over the hedonic tone view.

In light of the failure of these objections—and the fact that I have yet to find any better—I conclude that we should embrace the phenomenological thesis, but suspend judgment between the separate experience view and the hedonic tone view of hedonic states.

3. Evaluative Objections

This chapter concerns evaluative objections to BPH, i.e. arguments against BPH that rely on one or more evaluative premises. As one would expect, most of these objections directly target BPH's evaluative thesis:

Evaluative Thesis: The amount of welfare in a life (or life-segment) L, for its subject S, is determined by, and equal to, the sum of the magnitudes of all of S's pleasures during L minus the sum of the magnitudes of all of S's pains during L.

The structure of the chapter is as follows.

In §3.1, I address arguments for the claim that, contrary to the evaluative thesis, things other than pleasure and pain have intrinsic prudential value. In the process, I will develop and apply a general strategy for responding to evaluative objections against BPH, which I call "the undermining strategy".

In §3.2, I respond to the charge that the undermining strategy is epistemically self-defeating.

Finally, in §3.3, I rebut three evaluative objections that are not solely directed at the evaluative thesis. The first two are evaluative arguments against the calculation thesis. The third concerns both the evaluative thesis and the phenomenological thesis. The latter says that if the phenomenological thesis is true, then the evaluative thesis presents an implausibly alienating account of prudential value (and so BPH, which embraces both theses, is implausible as well). I will respond to all of these at least in part by applying the undermining strategy developed in §3.1 and defended in §3.2. For the second, I will also respond with an evaluative objection of my own.

3.1 Prudential Insight and the Undermining Strategy

3.1.1 Life Comparisons and the Strong Insight Hypothesis

Every form of prudential hedonism holds that only pleasures and pains have intrinsic prudential value. This claim has attracted many counterarguments, each attempting to establish that some things other than pleasure and pain can affect our welfare in themselves. In this section, I will develop a general strategy for rebutting these arguments, which will then provide a blueprint for my rebuttals to evaluative objections of other kinds.

The most compelling of these arguments employ a kind of thought experiment that I will call a “life comparison”. Life comparisons present the reader with a third-person comparison between two hypothetical lives, designed to elicit intuitions to the effect that one of the compared lives contains something of intrinsic prudential value (positive or negative) that the other lacks.

Life comparisons are often used by “objective list” theorists of welfare, who hold that there are multiple objective sources of intrinsic prudential value. Fletcher (2013: pp. 218-219) describes the method:

We should imagine two people with identical bundles of identical goods and then imagine some addition of some purported extra intrinsic welfare contributor or detractor. If we find it plausible to think that their welfare is thereby made unequal, this is evidence that we should expand the list to include this item.

Hooker (2015: pp. 23-24) puts this method into action, with autonomy as the candidate intrinsic prudential good:

We imagine two possible lives for someone as similar as possible except that one contains more autonomy and the other less. Then we ask which of these two

possible lives is more beneficial to the person who lives it. *The autonomous life seems better . . .* Since we have imagined that the two lives are as equal as possible in terms of the other elements of welfare, the best explanation of the superiority of the more autonomous life in this comparison is that autonomy is an element of welfare [i.e. has positive intrinsic prudential value].

From this and other identically structured life comparisons, Hooker concludes that not only autonomy, but also friendship, significant achievement, and important knowledge have intrinsic positive prudential value. Of course, if any of these conclusions are right, then the evaluative thesis is false. So this sort of life comparison presents one sort of evaluative objection to BPH.

Some other life comparisons, meanwhile, are used in arguments against specific theories of welfare. Dorsey (2011: 189) describes this method as a way of refuting prudential hedonism:

Imagine two lives that are precisely equal in pleasure. Imagine now, however, that the first life contains a greater degree of x than the second. Use x to mean whatever the purported non-pleasure good might be. For instance, let x be achieved goals. Intuitive judgment of two lives, identical in terms of pleasure, appears to lend weight to the suggestion that the successful life, rather than the unsuccessful life, is of greater welfare value.

Fletcher (2016: pp. 15-16) employs this same method with his variation on Nozick's (1974: pp. 42-45) experience machine scenario:

Trudy lives in New York. When not carrying on her groundbreaking research into stem cell treatment, she enjoys running marathons, working for a local charity, skiing, socialising with friends and spending time with her life and partner and her children. She also somehow finds time to pen highly successful, critically acclaimed novels. She enjoys great physical health and springs out of bed every morning full of joy and excitement.

Now meet Flora. When Flora was born she was attached to a machine that produces sensory stimulation that gives her very rich, vivid, and life-like experiences. She has the pleasurable experience of carrying on groundbreaking

research into stem cell treatment, of running marathons, of working for a local charity, skiing, socialising with friends and spending time with her life partner and their children. She also has the experience of writing highly successful and critically acclaimed novels. She is kept physically healthy by the machine and she also has the experience of springing out of bed every morning full of joy and excitement.

Let us stipulate that Trudy and Flora have lives of identical length and that their hedonic levels are identical at every moment of these lives . . . *Whatever* level of well-being one thinks is plausible for Trudy, it seems highly counterintuitive that Flora has that level of well-being also.

The counterintuitive claim in question is exactly the judgment that BPH endorses in this case. Trudy and Flora's lives are experientially identical, which makes them hedonically identical according to the phenomenological thesis; this, in turn, makes them equal in the amounts of pleasure and pain they contain according to the calculation thesis; and this, finally, makes them equal in welfare according to the evaluative thesis.

Lin (2016: p. 321) gives a similar version of the experience machine scenario (emphasis in original):

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

Do A and B contain the same total amount of welfare? I believe that many would join me in having the *comparison intuition*: A is at least somewhat higher in total welfare than B.²⁵

Like Fletcher, Lin stipulates that the compared lives (A and B) are experientially identical. So once again, the compared lives are hedonically identical according to the phenomenological thesis; which, in turn, makes them equal in the amounts of pleasure and pain they contain according to the calculation thesis; and this, finally, makes them equal in welfare according to the evaluative thesis. Hence BPH cannot accommodate Lin's "comparison intuition".

Finally, Velleman (1991: pp. 49-50) gives a similarly structured life comparison against additive theories of welfare, of which BPH is one:

Consider two different lives that you might have. One life begins in the depths but takes an upward trend: a childhood of deprivation, a troubled youth, struggles and setbacks in early adulthood, followed finally by success and satisfaction in middle age and a peaceful retirement. Another life begins at the heights but slides downhill: a blissful childhood and youth, precocious triumphs and rewards in early adulthood, followed by a midlife strewn with disasters that lead to misery in old age. Surely we can imagine two such lives as containing equal sums of momentary well-being. Your retirement is as blessed in one life as your childhood is in the other; your nonage is as blighted in one life as your dotage is in the other . . . To most people, I think, the former story would seem like a better life story – not, of course, in the sense that it makes for a better story in the telling or the hearing, but rather in the sense that it is the story of a better life. (Velleman 1991: 49-50.)

Here, Velleman stipulates that the two lives contain an equal "sum of momentary well-being". He takes no specific position as to what precisely goes into this sum. The point is to show that, in addition to momentary well-being, a life's overall shape or trajectory

²⁵ Other philosophers who have discussed a life comparison version of the experience machine include Crisp (2006: pp. 117-119), Hawkins (2016: pp. 361-363), and Pummer (2017: pp. 276-277).

makes an intrinsic contribution to its prudential value. The proffered evidence for this is that the story of the life with an upward-trending trajectory “would seem like a better life story” to most people than that of the life downward-trending life. If BPH is true, then this seeming must be mistaken, as according to the evaluative thesis, the amount of welfare in a life and a life’s sum of momentary well-being are one and the same.

Each of these passages contains an argument—what we may call a “life comparison argument”—that relies on an appeal to the reader’s intrinsic prudential intuition. This claim is central to all that follows in this section, so I will now explain just what I mean by intrinsic prudential intuition and how these arguments rely on appeal thereto.

Following Climenhaga (2015), I take it that S has an intuition that P just in case S feels inclined to believe that P, and S does not take that inclination to be the direct result of sense perception, testimony, memory, or inference. I further take it that one can have an intuition that P whether or not one already believes that P. I do not claim that this is the single true meaning of “intuition”, or that this is what philosophers always do, or should, mean when they use the term. I do claim that the objections I discuss in this chapter appeal to intuitions in this sense.

As I use the phrase, intrinsic prudential intuitions are intuitions about what has intrinsic prudential value. These include intuitions of the form “X has (lacks) intrinsic prudential value” and “X is good (bad) for a person in itself”, as well as intuitions with contents of the form “A is prudentially better than B”, where (given how A and B are

spelled out) this proposition either entails or is best explained by A being higher in intrinsic prudential value than B.

When I say that a philosopher appeals to intrinsic prudential intuition, I mean that they appeal to the fact that we have certain intrinsic prudential intuitions. I take it that for these philosophers, the mental event of having a certain intuition, at least for some people under some conditions, is evidence for the truth of the intuition's propositional content. The alternative would be to interpret them as asserting the propositional contents themselves. But if that were the case, these philosophers would not be giving arguments at all, but rather illustrations of points already arrived at. That interpretation is too uncharitable.

When I say that the arguments against BPH discussed in this chapter rely on appeal to intrinsic prudential intuition, I mean that in each argument, there is a premise for which this kind of appeal is the only support given. For example, consider the arguments given above by Hooker (two lives as identical as possible except with respect to achievement), Fletcher (the experience machine scenario featuring Trudy and Flora), Lin (the experience machine scenario featuring Adam and Bill), and Velleman (the upward- and downward-trending lives). Each one can be understood as having the following structure:

- P1. If Life A is higher in welfare than Life B, then Theory X is false.
- P2. Life A is higher in welfare than Life B.
- C. Theory X is false.

For any of the above life comparisons, we could fill in the blanks (substituting BPH for Theory X, since that is what we are concerned with here). For example:

- P1. If Trudy's life is higher in welfare than Flora's life, then BPH is false.
- P2. Trudy's life is higher in welfare than Flora's life.
- C. BPH is false.

This is a logically valid argument structure, and in every case, P1 will be true by definition. So the contentious premise will always be P2. The question, then, is what reason we are given to believe P2 is true.

Returning to the passages quoted above, we find Hooker saying that "the autonomous life seems better", Fletcher saying that "it seems highly counterintuitive" that Trudy's and Flora's lives would be equal in welfare, Lin referencing the "comparison intuition" that "A is at least somewhat higher in total welfare than B", and Velleman speculating that to most, the upward-trending life would seem like the story of a better life when compared to the downward-trending one. I interpret all of these assertions as appeals to intrinsic prudential intuition; that is, as appeals to a felt inclination on the part of the reader to believe the relevant proposition, where that inclination does not strike the reader as being the result of sense perception, memory, or inference, but rather as a sort of spontaneous intellectual response to the question being posed. And, crucially, these philosophers offer no other support for P2 (the reader will have to trust that I am not conveniently omitting anything). Hence my claim that these arguments rely on appeal to intrinsic prudential intuition.

For these arguments to succeed, given their reliance on appeal to intrinsic prudential intuition, two things must be true. First, people must actually have the intuitions in question. And second, the fact of people having these intuitions must give us sufficient reason to accept P2. I will not dispute the former. I don't know how widely

shared these intuitions are, and resting my defense on this question would turn it into an empirical project that I am not qualified to carry out. So, in what follows, I will just grant for the sake of argument that people generally have whatever intuitions are most convenient for my opponents. But I will dispute the second claim. Even if everyone had these intuitions, and we knew this to be the case, that would not give us adequate reason to accept P2. Or so I will now argue.

Though they do not say so explicitly, Hooker, Fletcher, Lin, and Velleman all suggest that the intuitions they appeal to give us evidence for their arguments' respective versions of P2. In other words, they suggest that our intrinsic prudential intuitions, at least about the given life comparisons, give us evidence for the truth of their contents. But these intuitions provide such evidence only if they reliably track the truth about intrinsic prudential value—reliably enough, that is, to make beliefs in their contents at least *prima facie* justified. (In what follows I will use the term “reliable” as shorthand for reliability in this sense.)

So, for these arguments to succeed, we must be justified in believing that our prudential intuitions are reliable. In other words, we must be justified in believing that we have what I will call “strong prudential insight”:

Strong Prudential Insight: A subject has strong prudential insight iff their intrinsic prudential intuitions reliably track the truth about what has (or lacks) intrinsic prudential value.

In summary: for the above evaluative objections to BPH (and others structured similarly) to succeed, we must be justified in accepting each argument's version of P2. For this, it must be legitimate to appeal to intrinsic prudential intuitions as evidence for the truth of

their contents. This, in turn, requires that we be justified in accepting an explanation (at a certain level) of why we have the specific intrinsic prudential intuitions that we do: that we have them because we have strong prudential insight.²⁶

Call this explanation the “strong insight hypothesis”. Now, I take this to be an uncontroversial fact about hypotheses: given two competing hypotheses about the same phenomena, and no decisive reason to favor one over the other, then we rationally ought to suspend judgment between them. So, to be justified in accepting the strong insight hypothesis (rather than suspending judgment about the relationship between our intrinsic prudential intuitions and the truth), we must have all-things-considered reason to accept it over incompatible alternative accounts of the intuitions in question. If there is such an alternative, and we do not have all-things-considered reason to reject it, then the mere availability of the alternative undermines our justification for believing that we have strong prudential insight.

This, then, will be my overarching strategy for rebutting evaluative objections to BPH: give an alternative, pro-hedonist hypothesis about our intrinsic prudential intuitions, and show that we are not justified in rejecting it in favor of the strong insight hypothesis. In what follows I will refer to this as “the undermining strategy”.²⁷

²⁶ I am taking internalism about epistemic justification for granted here. But the philosophers who appeal to evaluative intuition appear to assume some form of internalism as well. I am merely following their lead.

²⁷ Attentive readers will notice the similarity of this strategy to the one deployed in §2.3 and §2.4. The latter could just as well be called an “undermining strategy”, but for ease of exposition, I use that phrase in this chapter only.

3.1.2 The Undermining Strategy

The basic strategy I have just proposed is well-known. Every prudential hedonist that I am aware of employs some version of it. Sobel (2002: p. 244), an opponent of prudential hedonism in all forms, describes it well:

When a theory clashes this directly with central intuitions, the theory can regain plausibility by either (1) clarifying our intuition such that we come to see that when we properly understand the intuition, we see that it is not actually incompatible with the theory, or (2) explaining away the intuition by telling a story that undermines the credibility of the intuition...

The quantitative hedonist who conceives of pleasure as a sensation could adopt the second strategy of attempting to undermine our faith in the intuitions that run counter to the theory by telling a convincing story about the genesis of such intuitions that would explain why we have them while revealing them to be misleading in the cases in which they run counter to hedonism . . . Briefly the story that the quantitative hedonist would try to make convincing is that even our intuitions that run counter to hedonism can be explained by the truth of hedonism.

This is indeed the sort of approach I will pursue with my alternative hypothesis. But while past defenders of hedonism have most often used this sort of strategy to fend off evaluative objections one-by-one,²⁸ I will attempt to distinguish myself by giving a single pro-hedonist hypothesis that handles all evaluative objections sharing the basic structure of those given in the previous section, and that provides a foundation for responding to virtually any other objection relying on evaluative premises.

As Sobel says, the idea is to show that the intrinsic prudential intuitions that run counter to hedonism—such as the intuitions that Hooker et al seek to elicit with the life

²⁸ See for example Kawall (1999), Silverstein (2000), Baber (2008), Hewitt (2010), Feldman (2011), Barber (2011), and Weijers (2014) on the experience machine, and Feldman's (2006: pp. 108-131) and Crisp's (2006: pp. 111-125) piecemeal approaches to evaluative objections.

comparisons quoted above—can be (convincingly) explained by the truth of hedonism. So my hypothesis must assume the truth of BPH, and accomplish two further goals: predict our anti-hedonist intrinsic prudential intuitions, and avoid having any implausible implications. Here is what I propose:

The Pro-Hedonist Hypothesis: BPH is true, and we tend to correctly assign positive intrinsic prudential value to pleasure. But we also tend to mistakenly assign positive intrinsic prudential value to other things, just in case they meet these three conditions:

- (a) We observe them to be consistently conducive to net pleasure for people in general.
- (b) They are not categorically instrumental.
- (c) They are not activities that are usually done for some external end.

Our intrinsic prudential intuitions are the products of tacit inferences from features of the given scenarios plus these prior assignments of intrinsic value. (*Mutatis mutandis* for pain and negative intrinsic prudential value.)²⁹

This hypothesis suggests that we have a tendency to form beliefs, consciously or unconsciously, about what in the world is intrinsically harmful or beneficial to us, and that in doing so we begin (correctly) with pleasure and pain. Then, over a period of exposure to consistent correlations between hedonic states and the various non-hedonic things that we encounter in life, we begin to (incorrectly) assign intrinsic value to some of the things we observe to be consistently conducive to hedonic states. But only some. As we go, we filter out various candidate beliefs about intrinsic prudential value in a systematic way, arriving at plausible general principles that can receive widespread

²⁹ With the use of the phrase “net pleasure”, I do not mean to commit BPH to the view that pleasure and pain are commensurable opposites at the descriptive (non-evaluative) level. That said, the pro-hedonist hypothesis does require that people in general share a sense that something can bring on more pleasure than pain or *vice versa*.

assent across social contexts. And all this takes place more or less unconsciously, resulting in beliefs that are recalcitrant to revision even when we recognize the hedonic correlations that led to them in the first place.

Importantly, the pro-hedonist hypothesis does not predict that we will assign intrinsic prudential value to just anything that we observe to be consistently correlated with pleasure or pain for just anyone. Rather, it suggests that we assign such value only to things that we observe to be thusly correlated for people in general. Hence, for example, it does not predict that we will assign positive intrinsic prudential value to our own personal sources of pleasure when we know that our tastes are not widely shared. I consistently take great pleasure in listening to rather abrasive music; the pro-hedonist hypothesis does not predict that I will therefore come to believe that such music is good for people in itself.

Furthermore, the pro-hedonist hypothesis says that we assign positive intrinsic prudential value only to things that are not categorically instrumental, i.e. not defined wholly or partially by their instrumental properties. Hence it does not predict that we will assign such value to things like food, which is by definition a means to bodily sustenance; medicine, which is by definition a means to physical health; and money, which is by definition a means to acquire goods and services.

Finally, the pro-hedonist hypothesis says that we refrain from assigning positive intrinsic prudential value to activities that are usually done for some external end. The activities of eating, drinking, bathing, sex, and masturbation are almost universally associated with net pleasure, but the pro-hedonist hypothesis does not predict that we

will assign positive intrinsic prudential value to such activities, because we generally engage in them with some extrinsic goal in mind: nourishment in the case of food, cleanliness in the case of bathing, and net pleasure for all of the above.

In short, my hypothesis proposes that we tend to assign positive intrinsic prudential value to things that we strongly associate with net pleasure, provided that these things do not wear their instrumental natures on their sleeves, so to speak. This leaves things that are familiar enough to have strong hedonic associations, but abstract enough not to be obviously instrumental: things like autonomy, love, friendship, significant achievement, important knowledge, and positive life-trajectory. The same is true, *mutatis mutandis*, for pain and negative intrinsic prudential value: here the pro-hedonist hypothesis rules out the likes of debt and dental surgery as things to which we would assign intrinsic negative value, but leave things like ignorance, helplessness, and negative life-trajectory.

One might object that it is implausible for the pro-hedonist hypothesis to say that people in general have all these beliefs about intrinsic prudential value. I grant that this claim is speculative, and could perhaps be empirically disconfirmed. But I do not think it is implausible on its face. Here is why.

To begin with, beliefs about what is good and bad for people strike me as fairly ordinary, as does the general distinction between something's having value in itself and something's being valuable as a means to something else. And that is all the conceptual apparatus one needs to have beliefs about intrinsic prudential value. Moreover, I believe people in general are sufficiently acquainted with the likes of pleasure, autonomy, love,

friendship, achievement, knowledge, pain, ignorance, helplessness, and a person's life getting better or worse for them over time to have beliefs about such things.

But most importantly, recall that I am assuming, for the sake of argument, that people have whatever intrinsic prudential intuitions are most convenient for BPH's opponents. Presumably, the most convenient situation for my opponents would be one in which people in general have (or would have) anti-BPH intrinsic prudential intuitions about the relevant life comparisons. Now, for people to have these intuitions, and for these intuitions to be epistemically credible, people would need to be able to understand the cases being described and the basis on which they are to be compared. Someone could not have a credible intuition to the effect that Trudy's life is higher in welfare than Flora's without understanding the concept of personal welfare. And it stands to reason that a person who has such understanding would also have some beliefs about what sorts of things are good or bad for people.

Hence an assumption on which life comparison arguments rest becomes more dubious the more we resist the idea that people in general might have beliefs about intrinsic prudential value of the sort that the pro-hedonist hypothesis imputes to them. Anti-hedonists question this aspect of the pro-hedonist hypothesis at their own risk.

Earlier I set two goals for the pro-hedonist hypothesis. The first was to account for our anti-hedonist intrinsic prudential intuitions. Let's see how the hypothesis fares with the intuitions elicited by the life comparisons quoted earlier.

Hooker proposes a comparison between two lives that are as identical as possible, except that one contains more achievement. This (we assume) elicits the intuition that the

life with more achievement is higher in welfare. The pro-hedonist hypothesis accounts for this because, as noted, it predicts that we will assign intrinsic positive prudential value to achievement.

Fletcher and Lin each offer a life comparison version of the experience machine scenario, in which two people live lives that are experientially identical, one in the real world and the other in a simulation. These (we assume) elicit the intuition that the life spent in the real world is higher in welfare than the life spent hooked to the experience machine. The pro-hedonist hypothesis accounts for this as well. As noted, it predicts that we will assign positive intrinsic prudential value to things that the non-machine life has and the machine life lacks, such as friendship, autonomy, and significant achievement. It also predicts that we will assign negative intrinsic prudential value to things that the machine life has more of than the non-machine life, such as ignorance and helplessness (the person hooked to the machine being fundamentally ignorant about the circumstances of their lives, and completely dependent on the machine to carry on providing a facsimile of the real world).

Finally, Velleman compares two lives that have equal amounts of momentary well-being, but opposite welfare trajectories. This (we assume) elicits the intuition that the upward-trending life is higher in welfare than the downward-trending one. The pro-hedonist hypothesis accounts for this because, as noted, it predicts that we will assign positive intrinsic prudential value to an upward life-trajectory, and negative intrinsic prudential value to a downward life-trajectory.

Evidently, then, the intuitions elicited by these anti-hedonist life comparisons can be explained by the truth of BPH. For an anti-hedonist to give an argument that succeeds where these have failed, they will have to offer a comparison between two lives that differ with respect to some candidate intrinsic good that is not consistently conducive to net pleasure for people in general, or with respect to some candidate intrinsic bad that is not consistently conducive to net pain for people in general, or with respect to something that is categorically instrumental, or with respect to an activity usually done for the sake of an external end. And it will have to be the case that most people still find it intuitive that one of the compared lives is higher in welfare than the other. I cannot prove that there are no such comparisons, but I have not found any in the literature, and I am confident that none are forthcoming.

The second goal was for the pro-hedonist hypothesis to avoid implausible implications. I cannot prove that it has no such implications. I can, however, show that it avoids some implausible implications that befall the most prominent (and, to my knowledge, only) previous attempt at a wide-ranging psychological hypothesis in defense of prudential hedonism: that of Mill (2001: pp. 36-38).

In the passage I have in mind, Mill wants to show that hedonistic utilitarianism—i.e. utilitarianism founded on a hedonist theory of welfare—is not refuted by people having intrinsic desires for things other than pleasure. In typical hedonist fashion, he proposes that that these desires are best explained by the truth of hedonism. Mill's idea is that we come to desire these other things—he names virtue, money, power, and fame as

examples—only after, and precisely because, we have found them to be consistently conducive to pleasure. For instance, he says this about virtue (Mill 2001: p. 38):

Virtue, according to the utilitarian conception, is a good of this description. There was no original desire of it, or motive to it, save its conduciveness to pleasure, and especially to protection from pain. But through the association thus formed it may be felt a good in itself, and desired as such with as great intensity as any other good . . .

...Those who desire virtue for its own sake desire it either because the consciousness of it is a pleasure, or because the consciousness of being without it is a pain, or for both reasons united; as in truth the pleasure and pain seldom exist separately, but almost always together—the same person feeling pleasure in the degree of virtue attained, and pain in not having attained more. If one of these gave him no pleasure, and the other no pain, he would not love or desire virtue, or would desire it only for the other benefits which it might produce to himself or to persons whom he cared for.³⁰

Mill is accounting for desires here, not intuitions. But we can turn these remarks into an account of anti-hedonist intrinsic prudential intuitions just by adding the premise that the intuitions reflect our intrinsic desires (see Silverstein 2000: pp. 293-298). This account—call it the “Millian hypothesis”—does just as well as my own at explaining anti-hedonist intrinsic prudential intuitions like those elicited by the life comparisons discussed above.

However, the Millian hypothesis makes many other, much less plausible predictions about what intrinsic prudential intuitions we are likely to have. As noted earlier, Mill cites money, power, and fame as things that are sometimes desired for their own sake. He touts this as a fact that his hypothesis can explain. But if we repurpose Mill’s proposal as a hypothesis about intrinsic prudential intuitions, and not just intrinsic

³⁰ Similar accounts of the relationship between hedonic states and intrinsic desires were later proposed by Brandt (1979: pp. 95-98) and Railton (1989: pp. 167-171).

desires, then the hypothesis predicts that at least some of us will have intuitions to the effect that money, power, and fame have positive prudential value in themselves. Indeed, in the case of money, it predicts that virtually all of us would do so, money being consistently conducive to net pleasure for nearly everyone.

To see why this is a problem, consider another life comparison. Imagine two lives, one belonging to subject X and the other to subject Y, that are as identical as possible except that X has more money at every point in his life than Y does at the corresponding point in his life. Because the lives are otherwise as identical as possible, we must suppose that X's greater wealth does not produce an advantage for him in net pleasure or any other candidate prudential good. Given these stipulations, the Millian hypothesis predicts that most of us would have the intuition that X's life, the wealthier life, is higher in welfare. I would confidently predict otherwise.

This is not the only implausible implication of the Millian hypothesis. It generates similar predictions about every other thing that is consistently conducive to net pleasure for people in general, such as food, water, bathing, sex, and so on. Indeed, it predicts that, for any object O that is sufficiently conducive to net pleasure for you, and two hypothetical lives that are as similar as possible except that one contains more of O, you will have the intuition that the life with more of O is higher in welfare (provided that you understand and contemplate the comparison). And nothing in the Millian hypothesis precludes O from being something idiosyncratic. For example, if listening to the Beatles' *Abbey Road* is sufficiently conducive to net pleasure for me, then on the Millian hypothesis, I will have the intuition that a life—any life, not just my own—with more time spent listening to

Abbey Road will be higher in welfare, all else being equal. The Millian hypothesis allows that these sorts of idiosyncratic prudential intuitions could be ubiquitous. But again, I would confidently predict that most people do not have such intuitions.

My pro-hedonist hypothesis avoids all of these implications. Condition (a) says that we only tend to assign intrinsic positive prudential value to things that we observe to be conducive to net pleasure for people in general (*mutatis mutandis* for negative intrinsic prudential value and net pain); this rules out idiosyncratic intrinsic prudential intuitions as a common phenomenon. Conditions (b) and (c), meanwhile, rule out widespread assignments of intrinsic prudential value to things like money and food, by means already noted above.

So, while I cannot show that my hypothesis has no implausible implications, I can (and do) claim that it is an improvement on this score over the Millian hypothesis. Granted, this claim is based on speculation about what intrinsic prudential intuitions people are likely to have. Fortunately, the Millian hypothesis is itself compatible with BPH. If I am wrong, and people's intuitions are in line with what the Millian hypothesis predicts, this poses no problem for my larger project.

Let's review. I have given several examples of evaluative objections prudential hedonism, and shown that each makes an appeal to intrinsic prudential intuition. I have further shown that the success of this appeal, and therefore the success of these objections (and all others sharing an equivalent structure), depends on our being justified in accepting the strong insight hypothesis. We are justified in accepting this hypothesis only if there is no other account, incompatible with strong prudential insight and compatible

with BPH, that does equally well at accounting for our intrinsic prudential intuitions. But there is such an account: the one provided by my pro-hedonist hypothesis. I conclude that we are not justified in accepting the strong insight hypothesis—not by these arguments, anyway—and therefore that these evaluative objections to BPH, and all others sharing an equivalent structure, fail.³¹

³¹ There is at least one sort of objection to BPH's evaluative thesis to which the pro-hedonist hypothesis does not apply. It is suggested by the following passage from Mayerfeld (see also Hurka 2011: pp. 55-58):

Suppose some drug became available that gave people a joy as intense as the pain averted by anesthesia, and suppose that there were no drawbacks in the consumption of this drug. It seems quite clear to me that the provision of this drug would be less important than the administration of anesthesia . . . The moral badness of the suffering overshadows the moral goodness of the happiness. (Mayerfeld 1992: p. 133).

We can adapt this example to concern prudential rather than moral value. For this, we should ask whether it is better for a person, prudentially speaking, to receive anesthesia before an operation that would otherwise be very painful (supposing that the anesthesia will reduce your pain during the operation to zero), or to go through the procedure without anesthesia and then be given a drug that safely provides an amount of pleasure equal in magnitude to pain just endured (and that erases your memories of both the procedure and the subsequent drug-induced pleasure).

This scenario puts us in much the same position as the life comparisons discussed above. BPH's evaluative thesis has it that neither option in this case better for you than the other. But suppose that people in general have the intuition that the anesthesia route is prudentially preferable. If we are justified in supposing that these intuitions are rooted in a sort of prudential insight—this time into the amounts of intrinsic prudential value that various things contribute, relative to one another—then they will give us reason to reject BPH. But we are justified in supposing this only if there is no other account of these intuitions, incompatible with our having this kind of prudential insight and compatible with BPH, that does equally well at accounting for the intuitions in question.

The pro-hedonist hypothesis issues no prediction about how we will weigh the intrinsic prudential contributions of pleasure and pain, so it is of no use here. But we can come up with a different undermining hypothesis for cases like these. Here is one rough proposal.

3.2 Is the Undermining Strategy Self-Defeating?

The undermining strategy invites the charge that it is self-defeating, on the grounds that a lightly altered version of it could be used against BPH itself. For this, an opponent would only need to show that the case for BPH also relies on appeal to some intrinsic prudential intuitions, and then give an account of those intuitions that undermines their claim to being the product of prudential insight. Suppose, for example, that the case for BPH relied on contemplating a comparison between two lives that are identical except that one contains more pleasure. In that case, even if everyone had the intuition that the pleasurable life is higher in welfare, the anti-hedonist could simply respond with their own undermining hypothesis.

BPH is true, but we tend to mistakenly assign greater prudential weight, unit-for-unit, to pain than to pleasure. This is due to the confluence of various contingent psychological facts, such as: that we are generally capable of both more intense and longer-lasting episodes of pain than of pleasure (try to recall a physical pleasure as intense and long-lasting as the pain from a broken bone, or a period of general physical pleasure as intense or long-lasting as the general physical unease felt during a bout with the flu, or an emotional pleasure as intense and long-lasting as the emotional pain from heartbreak or the loss of a loved one); and that pain-inducing experiences are generally less subject to diminishing hedonic returns than pleasure-inducing ones (for example, a prolonged massage will tend to lose its pleasantness much more quickly than any form of physical torture will cease to be painful).

This leads us to be generally more averse to the prospect of a painful experience than we are attracted to the prospect of a pleasurable one, which, in turn, leads to the unconscious formation of false beliefs to the effect that pain is more bad for us than pleasure is good. Our intuitions about cases like the one adapted from Mayerfeld are the products of unconscious inference from these preexisting beliefs.

This is just another hypothesis. But it is coherent, consistent with BPH, and accounts for the intuitions in question, so it suffices to undermine arguments like the one derived from Mayerfeld above.

For instance, a desire theorist could account for pro-hedonist life comparison intuitions by proposing that only satisfactions of intrinsic desires have positive prudential value in themselves, and that we take the more pleasurable life in this comparison to be higher in welfare only because we so often have intrinsic desires for pleasure. On this account, we mistakenly assign intrinsic prudential value to pleasure because of its repeat association with the real source of welfare, intrinsic desire-satisfaction. In light of this alternative, we would not be justified in accepting that our intrinsic prudential intuitions about this comparison is the product of prudential insight.

This criticism misses the mark, because the case for BPH does not rely on any appeal to intrinsic prudential intuition. The only appeal to prudential intuition that I make on behalf of BPH in this dissertation comes in chapter 1, where I propose that it paints a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. There, I claimed that BPH assigns intrinsic prudential value to things that it is *prima facie* plausible to think have an impact on our welfare, and that it vindicates the commonsense idea that it is generally to a person's benefit to have mutually rewarding relationships, success at worthwhile activities, strong moral character, and freedom from sickness, injury, and subjugation by others. I was attempting to elicit prudential intuitions from the reader with this, but these need not be intuitions about what has intrinsic prudential value. They need only be intuitions to the effect that pleasure, pain, mutually rewarding relationships, success at worthwhile activities, strong moral character, and freedom from sickness, injury, and subjugation by others have instrumental prudential value under normal life circumstances.

In other words, at the outset of my case for BPH, we need only agree that the things listed here are closely tied to the real sources of intrinsic prudential value, whatever those may be. If the reader is willing to grant this, then no further appeal to prudential intuition, intrinsic or otherwise, is strictly necessary. My strategy from there is to argue that BPH survives various descriptive and evaluative objections, and finally, is preferable to its rivals because it is simpler (see chapter 4).

Critics are unlikely to be satisfied by this response. They will surely reply: even if you don't appeal to intrinsic prudential intuition at any point, another version of the undermining strategy can be used against the prudential intuitions you do appeal to—namely those underlying your claims about what counts as a recognizable picture of a life that is good for its subject under normal material and psychological circumstances.

This version of the self-defeat charge is more difficult to rebut. As I said earlier, the undermining strategy rests on a general principle about hypotheses: that given two competing hypotheses about the same phenomena, and no decisive reason to favor one over the other, we rationally ought to suspend judgment between them. If that's right, then the fact that I don't rely on appeal to intrinsic prudential intuition is irrelevant. All a critic would need to undermine the case for BPH is a plausible alternative hypothesis about the prudential intuitions I do appeal to, on which those intuitions lack any epistemically substantive connection to the truth about welfare. For example, they could offer a skeptical hypothesis on which no positive theory of welfare is correct, and our prudential intuitions are just byproducts of desires (for friendship, achievement, and so on) which are themselves byproducts of evolution. They would not even need to endorse

this hypothesis; the point would just be that something precisely analogous to my own undermining strategy can be used against me.

To resist this line of attack, I need a principled basis for the idea that my opponents' appeals to intrinsic prudential intuition can be undermined in the manner I proposed, while my own appeals to (not necessarily intrinsic) prudential intuition cannot. Fortunately, I have a principle in mind for this. It is the principle of phenomenal conservatism:

Phenomenal Conservatism (PC): If it seems to *S* that *p*, then, in absence of defeaters, *S* thereby has some degree of justification for believing that *p*. (Huemer 2007: p. 30.)

At first glance, this principle appears to be most congenial to my opponents. If PC is right, and the intrinsic prudential intuitions that anti-hedonists appeal to are seemings in the relevant sense, then (in the absence of defeaters) we are justified in accepting their anti-hedonist conclusions. The burden of proof would then be on me to supply defeaters—that is, positive reasons to believe that those intuitions are false—or admit defeat myself.

This appearance dissolves when we consider what it must mean for it to “seem” to *S* that *p*. According to Huemer (2001: p. 9), the author of this principle, the seemings referred to by PC are conscious, propositional mental states. So, when it seems to *S* that *p*, *S* is in a conscious mental state with propositional content *p*. But that's not all. On Huemer's view, seemings are also non-belief, non-inferential states. So, when it seems to *S* that *p*, *S*'s seeming that *p* is not a conscious instance of a preexisting belief that *p*, nor is it the product of *S* inferring *p* from any other proposition.

Without these latter stipulations, PC would grant *prima facie* justification to everything one happens to consciously believe, and to every conscious belief that one infers from preexisting beliefs. One could accept a version of PC without these restrictions, and if any anti-hedonists are so inclined, then they will not be convinced by what remains of my defense of the undermining strategy. But I think this would make for an overly permissive epistemic principle. So, in what follows, I will adopt Huemer's understanding of "it seems to *S* that *p*".

Now notice that, on the pro-hedonist hypothesis, our anti-hedonist intrinsic prudential intuitions are not seemings in the relevant sense, nor are they based on such seemings. Instead, they are the products of tacit inferences from preexisting beliefs about what is intrinsically good for us. These beliefs, meanwhile, are not grounded in seemings with the corresponding propositional contents. According to the pro-hedonist hypothesis, the only intrinsic prudential intuitions that could count as seemings in the relevant sense are those that favor prudential hedonism. In other words, on my hypothesis, it never seems to us (in the relevant sense) that anything other than pleasure is good for us in itself, or that anything other than pain is bad for us in itself. All other intrinsic prudential intuitions are either conscious instances of unconsciously-formed beliefs, or the products of inference from such beliefs, and are therefore not the kinds of mental states to which PC grants *prima facie* justification.

What about the prudential intuitions underlying my claims in chapter 1 about pictures of lives that good for their subjects? If PC and the pro-hedonist hypothesis are both right, can these intuitions provide even initial justification for taking BPH seriously?

I believe so. On the view I am sketching, seemings confer *prima facie* justification on the beliefs that pleasure and pain have prudential value. As long as this justification can extend to other propositions via valid inference, we can become justified in believing that the other items mentioned in my sketch of a life that is good for its subject—mutually rewarding relationships, success at worthwhile activities, and so on—are at least instrumentally prudentially valuable under normal life circumstances. We would only need to recognize the ways that these things are typically correlated with pleasure and pain, and infer accordingly. And it is consistent with both PC and the pro-hedonist hypothesis to suppose that this is what is happening (subconsciously) when we find that some theory paints a recognizable picture of a life that is good for its subject under normal material and psychological circumstances.

Admittedly, my line of reasoning has become somewhat convoluted here. So let's review once more. I have argued that evaluative objections to BPH rely on appeal to intrinsic prudential intuition, and I have given a pro-hedonist hypothesis to undermine these objections. My hypothetical critics claim that this approach is self-defeating, because the case for BPH must also rely on appeal to intrinsic prudential intuition, and so an analogous strategy could be used against it. I reply that the case for BPH does not rely on intrinsic prudential intuition. However, I admit that it relies on appeal to prudential intuition of a sort. For my overarching argument in this dissertation to get off the ground, the reader must share certain intuitions about what counts as a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. The hypothetical critic responds that an analogous version of my undermining strategy

could be used against these intuitions. To rebut this charge of self-defeat, I need a principled basis for the claim that the undermining strategy works against my opponents' appeals to intuition, but not my own. For this, I propose PC. If both PC and the pro-hedonist hypothesis are right, then my appeals to prudential intuition are ultimately based in the seemings that confer *prima facie* justification, while those of my opponents are not.

I expect that critics—perhaps somewhat exasperated at this point—will reply that I have given no positive reason to believe that either PC or the pro-hedonist hypothesis are true. This is correct. I have only floated them as possibilities. But the mere fact that this combination is available to the basic prudential hedonist, and is not clearly implausible, means that opponents must do more than give arguments relying on appeal to intrinsic prudential intuition. They need to give us some reason, independent of these intuitions, to believe that we have strong prudential insight. One way to do this would be to find a direct route to the truth about what has prudential value in itself that circumvents appeal to intuition altogether. Alternatively, critics could find some implausible implication of the pro-hedonist hypothesis that has escaped my notice, or give an argument against PC. In any case, the project of showing that things other than hedonic states have intrinsic prudential value will be more complicated than many have thought.

3.3 More Evaluative Objections

In §3.1, I developed the undermining strategy in the context of arguments to the effect that some things other than hedonic states have intrinsic prudential value. I will now apply versions of this strategy to other sorts of evaluative objections. In §3.3.1, I will

discuss two evaluative objections to the calculation thesis. And in §3.3.2, I will discuss an objection that concerns both the evaluative thesis and the phenomenological thesis. In this last section, in addition to applying a version of the undermining strategy, I will respond with an evaluative objection of my own.

3.3.1 Evaluative Arguments Against the Calculation Thesis

Roger Crisp, a fellow prudential hedonist, rejects the calculation thesis:

Calculation Thesis: All pleasures have magnitudes that are quantifiable (in principle, using cardinal numbers) and fully commensurable; the same is true of pains.

He claims that there are discontinuities in the hedonic calculus, in the form of quality distinctions among pleasures. These quality distinctions are such that for two pleasures of disparate quality, X and Y, any amount of lower-quality pleasure X will never contain as much pleasure as higher-quality pleasure Y. Crisp argues for this on evaluative grounds, using the following thought experiment:

Haydn and the Oyster. You are a soul in heaven waiting to be allocated a life on Earth. It is late Friday afternoon, and you watch anxiously as the supply of available lives dwindles. When your turn comes, the angel in charge offers you a choice between two lives, that of the composer Joseph Haydn and that of an oyster. Besides composing some wonderful music and influencing the evolution of the symphony, Haydn will meet with success and honour in his own lifetime, be cheerful and popular, travel, and gain much enjoyment from field sports. The oyster's life is far less exciting. Though this is rather a sophisticated oyster, its life will consist only of mild sensual pleasure, rather like that experienced by humans when floating very drunk in a warm bath. When you request the life of Haydn, the angel sighs, 'I'll never get rid of this oyster life. It's been hanging around for ages. Look, I'll offer you a special deal. Haydn will die at the age of seventy-seven. But I'll make the oyster life as long as you like'. (Crisp 2006: p. 112)

The idea here is that no matter how long the oyster life lasts, it will never be as high in welfare as the life of Haydn. This poses a threat to prudential hedonism, because a long

enough oyster life will have to contain more pleasure than Haydn's life. Unless, that is, we posit quality distinctions among pleasures. Then we can accept that Haydn's life is higher in welfare than the oyster's, no matter how long the latter lasts, while affirming that this is so precisely because Haydn's life is higher in total pleasure. After all, if Haydn's life contains higher-quality pleasures (from achievement, success, relationships, and so on), then no amount of mild sensual pleasure will ever be able to match it, let alone exceed it. But if there are such quality distinctions among pleasures, then not all pleasure magnitudes are commensurable with one another, which means the calculation thesis is false, and so BPH is false.

This argument is similar to the ones discussed in §3.1, in that it presents a comparison between two hypothetical lives that are described so as to elicit an intuition to the effect that one life has something of intrinsic prudential value that the other lacks. In this case, however, that thing is not something that is present in one life and wholly absent from the other; instead, it is something that exists at a certain level of quality in one life and not the other. Both lives contains lots of pleasure, but Haydn's life has pleasures of higher quality.

Despite this difference, Crisp's argument can be rebutted in much the same way as the objections discussed in §3.1. It shares the same basic structure as those arguments:

- P1. If Life A is higher in welfare than Life B, then Theory X is false.
- P2. Life A is higher in welfare than Life B.
- C. Theory X is false.

In this case, we would fill in the blanks like so:

- P1. If Haydn's life is higher in welfare than the oyster life, then BPH is false.
- P2. Haydn's life is higher in welfare than the oyster life.
- C. BPH is false.

Once again, the crucial premise is P2. And P2 is being supported by nothing but an appeal to intrinsic prudential intuition. Hence it relies on the assumption of strong prudential insight. All we need to undermine it is a plausible alternative hypothesis that is compatible with BPH.

Can the pro-hedonist hypothesis do the job? If we stipulate that the two lives are equal in net pleasure, then yes: the pro-hedonist hypothesis predicts the intuition that Haydn's life is higher in welfare than the oyster's. This is because the former contains the likes of achievement and (presumably) friendship—things strongly correlated with net pleasure for people in general that are neither activities usually done for some external end nor instrumental by definition—which the oyster life lacks completely.

However, this case is not quite so simple. Crisp adds the wrinkle that the oyster's life may last long enough to be indefinitely higher in net pleasure than Haydn's. The pro-hedonist hypothesis, as given earlier, makes no firm prediction about comparisons structured this way, because it tells us nothing about how we will weigh different purported sources of intrinsic prudential value, such as achievement and pleasure, against one another. But we can handle this with just a bit more psychological speculation.

Here's one way that might go. Crisp describes the oyster life as consisting of an indefinitely long period of "mild sensual pleasure, rather like that experienced by humans when floating very drunk in a warm bath." It could be that, when we are subconsciously forming judgments about the amounts of welfare in compared lives, we apply a

“monotony threshold”, such that there is only so high in welfare a life can be when its pleasures are all taken in the same sort of object or experience. To exceed this level of welfare, the life must be more varied in its sources of pleasure. The psychological roots of this could reside in the fact that monotony is strongly correlated with unpleasant experience. After all, as pleasant as a drunken bath can be, there inevitably comes a time when you want to sober up and dry off.

Again, this is only a psychological hypothesis. But it is consistent with BPH being true, accounts for the intuition in question, and has no implausible implications that I can see. If my assessment of it is right, then it establishes that we are not justified in accepting Crisp’s version of P2. So his argument fails, and the calculation thesis is safe.

Crisp’s example places the focus on allegedly higher-quality pleasures, such as those of achievement. Another approach to arguing for quality distinctions among pleasures is to argue for the qualitative inferiority of immoral pleasures. Feldman (2006: p. 39) helpfully provides a vivid example:

Suppose some terrorist really hates children. Suppose he sets off a bomb at a playground, and then watches the news on TV. When he sees the suffering children choking and gasping and bleeding, and learns of the many injuries and deaths, this terrorist feels a thrill of pleasure. His pleasure is caused by the misery of his victims.

Suppose the terrorist does this many times over, and each time is delighted by the fruits of his labor. Suppose at the same time that his life is not filled with counterbalancing pains.

For the sake of precision, we should imagine this life in contrast with a life equal in net pleasure, with pleasures all taken in morally neutral activities. BPH implies that these lives would be equal in welfare. But we may find it intuitive that the terrorist’s life is lower

in welfare than its counterpart. At the same time, we may find it intuitive that the terrorist's life is higher in welfare than a life with no pleasure, friendship, achievement, etc. at all, suggesting that immoral pleasures are not prudentially worthless. If these intuitions are correct, then calculating the amount of positive prudential value in a life is not simply a matter of adding up pleasure magnitudes, meaning that the calculation thesis is false, and so BPH is false.

Fortunately, the pro-hedonist hypothesis can account for these intuitions. Immorality is strongly correlated with net pain—via the pangs of conscience and social censure—for people in general (the occasional sadist or unrepentant murderer notwithstanding); hence the pro-hedonist hypothesis predicts that we will assign it negative intrinsic prudential value, albeit perhaps only subconsciously. This negative value partially offsets the positive value provided by the terrorist's pleasure, which accounts for the intuition that the terrorist's life is lower in welfare than one that is equal in net pleasure but morally sound. At the same time, it may be that the negative prudential value we assign to the terrorist's immorality is lower than the positive prudential value that we assign to his pleasure (provided that the pleasure is stipulated to be sufficiently intense—and Feldman does describe the terrorist as thrilled and delighted). This would explain why we have the intuition that the terrorist's life is better than a prudentially neutral one.

What if we allowed the terrorist's life to be indefinitely long, as Crisp did with the oyster? BPH would then imply that the terrorist's life is higher in welfare than its morally neutral counterpart. Once again, I would speculate that there is, in our minds, a monotony

threshold, such that there is only so good the terrorist's life can be when his pleasures all come from *schadenfreude*.

3.3.2 The Resonance Constraint

In chapter 2, I argued for the phenomenological thesis:

Phenomenological Thesis: All pleasures are mental states that feel the same way in some respect, and it is this common feeling that makes them count as pleasures; the same is true of pains.

This thesis implies that feeling pleasure or pain is just a matter of tokening a certain phenomenal property. Contrary to the attitudinal theories of hedonic states canvassed in the previous chapter, the phenomenological thesis allows that you can be in a hedonic state without holding any attitude toward any part of your present state of consciousness. For example, it implies that you can feel pleasure without wanting to feel what you are feeling. Furthermore, the phenomenological thesis allows that you can have a disfavorable attitude toward an occurrent pleasure, or a favorable attitude toward an occurrent pain. It implies that you can feel pleasure, have no favorable attitude toward that pleasure, and strongly want that pleasure not to be occurring. You can also feel pain, have no disfavorable attitude toward that pain, and strongly want it to be occurring. And these desires can be intrinsic, i.e. directed entirely at how those particular hedonic states feel.

The evaluative thesis, meanwhile, says that all pleasures have intrinsic positive prudential value, and all pains have negative intrinsic prudential value. The conjunction of this and the phenomenological thesis implies that something can have intrinsic prudential value for you, positive or negative, irrespective of your attitudes toward it (if

indeed you have any). More specifically, it implies that each of your pleasures is intrinsically prudentially good for you even when you have no attitude, or an intrinsically disfavorable attitude, toward it; and that each of your pains is intrinsically prudentially bad for you even when you have no attitude, or an intrinsically favorable attitude, toward it.

Moreover, the phenomenological and evaluative theses together imply that your attitudes, in themselves, make no difference to how a given hedonic state affects your welfare. According to BPH, two hedonic states of equal intensity and duration are of equal intrinsic prudential value, even if you strongly favor one and strongly disfavor the other. In short (and less formally), your pleasures are just as intrinsically good for you whether you like them or not, and your pains are just as intrinsically bad for you whether you like them or not.

In this way, BPH runs afoul of what has been called the “resonance constraint” on theories of welfare (or theories of value more generally).³² There is disagreement about exactly how to formulate this constraint, but the basic idea is set out in an oft-cited passage from Railton (1986: p. 9):

Is it true that all normative judgments must find an internal resonance in those to whom they are applied? While I do not find this thesis convincing as a claim about

³² Dorsey (2011: p. 185) and Bramble (2016: p. 85) use the exact phrase “resonance constraint”. The use of the term “resonance” in this way appears to originate with Railton (1986: p. 9). In that same paper, Railton uses the phrase “internalist constraint” in much the way Dorsey and Bramble use “resonance constraint”. Rosati (1996: p. 303) also defends what she calls an “internalist constraint”; in this connection she cites the aforementioned Railton (1986) as well as Darwall (1983: pp. 54-55), from whom she borrows the phrase “existence internalism”, which is the view that “there is a necessary connection between motivation and normative status” (Rosati 1996: p. 297). See also Velleman (1998) and Hawkins (forthcoming). I prefer the language of resonance over that of internalism here because “internalism” is used for so many different things in philosophy.

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

Railton proposes a that in order for something to be intrinsically good for you, it must have an "internal resonance" for you, meaning that it must be something you find compelling or attractive—if not under the actual circumstances, then at least under some ideal circumstances. If this is right, then it serves as a constraint on theories of welfare: if a theory allows that something can be intrinsically good for you without "resonating" for you in this way, then it gives "an intolerably alienated conception of someone's good", and is therefore to be rejected.

Thus Railton provides one example of a "resonance constraint". And it is a useful example, because although it is about as forgiving as such a constraint could be—it allows that something can be good for you even if you don't find it compelling or attractive under any of your actual circumstance—BPH still fails to meet it. As long as "internal resonance" is understood to require some actual or counterfactual attitude, motivation, or indeed anything other than phenomenology, BPH will allow that something—namely, pleasure—can be intrinsically good for you without resonating. If there is a corresponding constraint on the negative side, to the effect that something can be intrinsically bad for you only if you find it non-compelling or unattractive (at least under some ideal circumstances), then BPH will fail to meet that as well.

The question, then, is what reason we have to accept the resonance constraint. It is not self-evident, or a logical truth, so we will need to see it applied to examples. The starkest sort of example would be a life comparison structured like those discussed in §3.1. For this, we could imagine two lives that are as identical as possible with respect to everything that might be intrinsically good for us (without necessarily being resonant in the relevant sense); pleasure, achievement, and friendship might be examples. Then, we could imagine that in one life all of these things are resonant, whereas in the other life, none of them are. We would then have a life comparison argument for the resonance constraint, with a familiar structure:

P1. If the resonant life is higher in welfare than the non-resonant life, then theories that reject the resonance constraint (such as BPH) are false.

P2. The resonant life is higher in welfare than the non-resonant life.

C. Theories that reject the resonance constraint (such as BPH) are false.

I suspect that whether people have the intuition that the resonant life is higher in welfare than the non-resonant life will depend on what exactly resonance entails. Railton says that “what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware.” This leaves open the possibility that something can be resonant for you even if you don’t find it compelling or attractive under the actual circumstances in which you have it. In that case, the comparison could be between two people who find nothing in their lives compelling or attractive. The difference would be that one of them would find various things in his life compelling or attractive if he were rational and aware, whereas the other

would not. And yet the two lives would still be equal in pleasure, friendship, achievement, and so on. I find this sort of example difficult to imagine. Consequently, the claim that people in general would have the intuition that the resonant life is higher in welfare in this comparison strikes me as doubtful.

On the other hand, suppose we imagine that the person living the resonant life actually does find their pleasures, achievements, friendships, etc. compelling and attractive. In that case, it strikes me as much more probable that people in general would have the intuition that the resonant life is higher in welfare. But then this is precisely what the pro-hedonist hypothesis would predict. This is because the resonant life will be higher in something that is strongly correlated with net pleasure for people in general, without being instrumental by definition or an activity usually done for some external end: satisfaction of intrinsic desires. In this case, then, the standard undermining strategy works.

But suppose that some proponent of the resonance constraint utterly rejected the legitimacy of the undermining strategy. In that case, I would fight fire with fire, and give an intuition-based evaluative objection to the constraint itself. Here is that objection.

Recall that resonance must involve some non-phenomenological element. Otherwise, there is not necessarily any conflict between the resonance constraint and BPH. Given the connection between phenomenology and awareness (see §2.2.1), it follows that it is at least logically possible for the experience of possessing a resonating good to be phenomenally identical to, and therefore subjectively indistinguishable from, the experience of possessing that same thing without resonance.

Before I explain why this is troublesome for the resonance constraint, here is an illustration. Suppose that one way for something to resonate for you is for you to have a favorable attitude toward it, for its own sake, while you have it. Further suppose that part of what it means to have a favorable attitude toward something for its own sake while you have it is to be disposed, under certain conditions, to act so as to keep it. A behavioral disposition is not a phenomenal property, of course, so this would mean that resonance is at least partly non-phenomenal. It follows that it is logically possible for you to have something in a resonant way at t_1 , and to have that same thing in a non-resonant way at t_2 , without there being any difference between your conscious experience at t_1 and your conscious experience at t_2 .

For example, suppose you experience the taste of root beer at both t_1 and t_2 , but you only have a resonance-producing favorable attitude toward that taste at t_1 . It is nonetheless possible for your total conscious experience at t_1 to be exactly the same as it is at t_2 . Perhaps your experience at t_1 includes some phenomenology associated with the favorable attitude toward the taste of root beer; if so, then this too would be replicated at t_2 , even though at t_2 the attitude is absent. Again, so long as having the favorable attitude in question is not simply a matter of tokening a certain phenomenal property, this will remain a logical possibility. The upshot is that you are having a resonating experience at t_1 —an experience that meets the resonance constraint, and is therefore a *prima facie* candidate for being intrinsically good for you—and a non-resonating experience at t_2 , and to you these feel absolutely identical.

If there is an equivalent to the resonance constraint on the negative side—it would certainly be a puzzling asymmetry if there weren't—then the same point would apply, *mutatis mutandis*. Let's call the corresponding phenomenon "anti-resonance", and suppose that it involves having a disfavorable attitude toward something, for its own sake, while you have it. Now, for example, suppose you experience an intense pinching sensation at both t_1 and t_2 , but you only have an anti-resonance-producing disfavorable attitude toward that sensation at t_1 . It is nonetheless possible for your total conscious experience at t_2 to be exactly the same as it is at t_1 . Perhaps your experience at t_1 includes some phenomenology associated with the disfavorable attitude toward the pinching sensation; if so, then this too is replicated at t_2 , even though the attitude is absent. The upshot is that you are having an anti-resonating experience at t_1 —an experience that meets the anti-resonance constraint, and is therefore a *prima facie* candidate for being intrinsically bad for you—and a non-anti-resonating experience at t_2 , but to you they feel absolutely identical.

It follows that if the resonance constraint is right, then it is possible for you to have an experience that is subjectively identical to a euphoric, resonant pleasure, without that experience actually being resonant for you, and therefore without it actually being intrinsically good for you. It also follows that it is possible for you to have an experience that is subjectively identical to an agonizing, anti-resonant pain, without that experience actually being anti-resonant for you, and therefore without it actually being intrinsically bad for you. To see how these implications are counterintuitive, consider the following scenario:

You are in the hospital for an operation that will require you to be placed in a medically induced coma. Before the operation, you are told (correctly) that thanks to advances in brain-stimulation technology, you can now choose what, if anything, you would like to feel while on the operating table. You have three options, nicknamed “heaven”, “hell”, and “limbo”.

On the “heaven” option, you will feel nothing but euphoric emotional pleasure, as if you are simply overjoyed to be alive. On the “hell” option, you will feel nothing but agonizing emotional pain. Finally, on the “limbo” option, you will have no qualitative experience whatsoever.

All three options begin with a hedonically neutral procedure that renders you temporarily unable to form favorable or unfavorable attitudes. For the duration of the operation, you will have no desires, preferences, likes, or dislikes. This will have no effect on what, if anything, you feel. If you choose the heaven or hell options, you will still feel exactly as though you have favorable or unfavorable attitudes toward various aspects of your situation, respectively.

Finally, upon completion of the operation, you will have no memory of whatever experiences you had (if any).

The resonance and anti-resonance constraints together imply that, as far as your welfare is concerned, your choice among these three options would be objectively arbitrary. All three would have the same effect on the goodness of your life for you, which is to say no effect at all. Without the capacity to form attitudes, the pleasures of the heaven option could not resonate for you, and the pains of the hell option could not anti-resonate for you. This makes absolutely no difference to how they would feel—they would seem just as heavenly or hellish either way. But the effect of either option on your welfare would be nil, just as it would be if you chose limbo.

Now, suppose that instead of being presented with these options yourself, a hospital employee, a stranger to you, is given the task of choosing among them on your behalf. You are never even informed that these options exist. If the resonance and anti-

resonance constraints are right, then considerations of your welfare cannot give this employee any objective reason to choose one option over another. If there is any moral reason for this employee to choose the heaven option for you, or at least refrain from choosing the hell option, this reason cannot have anything to do with harming or benefitting you. Each option harms and benefits you exactly the same, which is not at all.

It strikes me as intuitively obvious that there is a reason, grounded in your welfare, for you to choose heaven rather than hell or limbo in the first case, and for the employee to choose heaven rather than hell or limbo for you in the second. I expect that the reader will agree. If so, then we can agree that the resonance constraint leads to deeply counterintuitive conclusions about this scenario. (If the reader does not agree, then I don't know what more to say.)

This argument relies on appeal to intrinsic prudential intuition, and so is, in my estimation, vulnerable to a version of the undermining strategy. But if a proponent of the resonance constraint responded this way, they would be granting the legitimacy of this strategy, which could then be used against arguments for the resonance constraint itself.

3.4 Conclusion

I know of no evaluative objection to BPH that does not ultimately depend on our being justified in believing that we have strong prudential insight. But the claim of our having such insight is undermined by the availability, and plausibility, of the pro-hedonist hypothesis. I conclude that all such objections leave BPH unscathed.

So far, then, I have argued that BPH is fundamentally appealing, and that it survives a battery of descriptive and evaluative objections that many have thought

decisive. In the following chapter, I will attempt to show that in addition to being defensible in these ways, BPH enjoys an advantage over its rivals: its simplicity.

4. Simplicity

Thus far, my discussion of basic prudential hedonism (BPH) has been entirely defensive. In chapter 1, I proposed that BPH is one among several theories of welfare that paints a recognizable picture of a life that is good for its subject under normal material and psychological circumstances; in chapters 2 and 3, I argued that BPH survives various descriptive and evaluative objections. If these arguments have been successful, then I have fought opponents of BPH to a standstill.

I would like to argue that BPH is not only defensible, but is in fact the best theory of welfare on offer. Unfortunately, the options for this are severely limited. Any positive argument for BPH based on appeal to prudential intuition would be vulnerable to the same sort of undermining strategy I used in defense of BPH in chapter 3. (This is why I do not regard my intuitive argument against the “resonance constraint” in §3.3.3 to be especially persuasive). And aside from unrestricted actualist desire theories (see §1.2.2), none of BPH’s rivals are logically inconsistent or otherwise incoherent in any way that I can discern. However, BPH does appear to have one distinct advantage over its main competitors: its simplicity.

I suspect that the apparent simplicity of prudential hedonism is a large part of what attracts the few adherents it has. This is certainly true in my case. Yet hedonists rarely cite simplicity as an advantage of their view explicitly. Moen (2016: p. 278, emphases mine) is an exception:

...if the suggested non-hedonic intrinsic values are potentially explainable by appeal to just pleasure and pain . . . then—*by appeal to Occam’s razor*—we have at least a *pro tanto* reason to resist the introduction of any further intrinsic values and

disvalues. It is *ontologically more costly* to posit a plurality of intrinsic values and disvalues, so in case all values admit of explanation by reference to a single intrinsic value and a single intrinsic disvalue, *we have reason to reject more complicated accounts*.

But that is all he says on the matter. Moen does not explain how theories of welfare that introduce non-hedonic intrinsic values are more complex, or ontologically costly, in a way that gives us reason to reject them. This is understandable. The philosophical literature on theoretical simplicity, which exists almost entirely outside of value theory, is large and vexing. Meanwhile, many readers may already share an intuitive sense of which theories of welfare are simpler and why this simplicity is desirable, making further discussion of these things both philosophically risky and rhetorically unnecessary. But given that, in my estimation, the positive case for accepting BPH as the best all-things-considered theory of welfare rests on its purported simplicity, closer examination of these points is in order.

My exploration of this topic will not be very ambitious. I will not attempt to settle any fundamental philosophical questions about simplicity. What I will do is outline an argument for the claim that the simplicity of BPH gives us epistemic—not merely aesthetic or pragmatic—reason to accept it. The reason I propose is that where all else is epistemically equal, simpler theories are more likely to be true. This argument relies on some contentious assumptions, full defense of which would take us beyond the scope of this dissertation. I will note these assumptions when they appear, and try to explain why I find them reasonable. My main contribution here will be to provide a clearer sense of what must be true for this sort of argument to work, especially in defense of a theory of

welfare; and thereby to highlight the points that stand most in need of further support for this approach to be made wholly convincing.

The structure of the chapter is as follows. In §4.1, I propose a specific criterion of simplicity, and explain how—with the help of some auxiliary assumptions—a theory being simpler by this standard gives us an epistemic reason to accept it, all else being epistemically equal. In §4.2, I argue that BPH is simpler than its rivals by this standard, and therefore—given the assumptions discussed in the previous section—is more likely to be true. Finally, in §4.3, I address some challenges to the position sketched in the previous sections.

4.1 Syntactic Simplicity and Epistemic Probability

4.1.1 A Syntactic Criterion of Theoretical Simplicity

It is infamously difficult to explain what makes one theory simpler than another, and why, all else being equal, simplicity is good. As Sober (2001: pp. 14-15) points out, seeking simplicity is always matter of minimizing the number of something; but what is being minimized in the name of simplicity can differ from one philosopher or context to another. There are, at minimum, two senses in which one theory may be simpler than another: an ontological sense, in which the simpler theory posits the existence of fewer entities, causes, or properties; and a syntactic sense, in which the simpler theory

minimizes the number of fundamental propositions it contains.³³ The criterion of simplicity I propose is a syntactic one:

Simplicity: One theory, T_1 , is simpler than another, T_2 , if and only if T_1 contains fewer fundamental claims than T_2 , where a claim is fundamental iff it is:

- (a) A non-vacuously quantified first-order proposition containing no singular terms.
- (b) Not logically equivalent to a conjunction of non-vacuously quantified first-order propositions containing no singular terms.
- (c) Not entailed by other propositions within the theory.

Fundamental claims, by this definition, are claims like “all pleasures have positive intrinsic prudential value” and “all frustrations of informed desires have negative intrinsic prudential value”.

Non-fundamental claims would include propositions to the effect that some token thing, such as a particular pleasure, has positive intrinsic prudential value. Such propositions would be unquantified, and contain at least one singular term referring to the token pleasure. If these non-quantified propositions were fundamental, then every theory would contain an infinite number of fundamental claims. It may be that one infinite set of fundamental claims can be smaller in size than another infinite set, by dint of some mathematical facts beyond my comprehension, but I suggest we avoid this complication altogether by counting only quantified claims as fundamental.

Notably, claims such as “all and only pleasures have positive intrinsic prudential value” are also non-fundamental. This is because claims of this form are universally

³³ To my knowledge, the main exponent of the syntactic approach to simplicity has been Nelson Goodman, though his view is that to assess a theory’s simplicity is to assess the structure of its system of primitive predicates. For a relatively accessible introduction see Goodman (1958).

quantified biconditionals, which are logically equivalent to conjunctions of universally quantified conditionals. Let P stand for the predicate “is a pleasure”, and V stand for the predicate “has positive intrinsic prudential value”. The proposition “all and only pleasures have positive intrinsic prudential value” can be written as follows:

$$(\forall x)(Px \equiv Vx)$$

This is one non-vacuously quantified first-order proposition. But it is a biconditional, and is therefore logically equivalent to a conjunction:

$$(\forall x)(Px \supset Vx) \ \& \ (\forall x)(Vx \supset Px)$$

Condition (b) in the above definition effectively demands that the biconditional be expressed as a conjunction, while (a) demands that each conjunct be counted as a separate fundamental claim (as the conjunction itself is not a quantified proposition).

This is important, because without these demands, a theory that assigns intrinsic prudential value to fewer things would not *ipso facto* be any simpler than a theory that assigns such value to more. As we have just seen, it takes two fundamental claims to express the idea that exactly one type of thing, pleasure, has positive intrinsic prudential value. What about a theory that assigns such value to two types of things—pleasure and autonomy? Let A stand for the predicate “is an instance of autonomy”. Without condition (b), we could express this part of the theory in just two fundamental claims also:

$$(\forall x)[(Px \vee Ax) \supset Vx], \ (\forall x)[Vx \supset (Px \vee Ax)]$$

But notice that $(\forall x)[(Px \vee Ax) \supset Vx]$ is also equivalent to a conjunction of two non-vacuously quantified first-order propositions:

$$(\forall x)(Px \supset Vx) \ \& \ (\forall x)(Ax \supset Vx)$$

Hence the criterion of simplicity I have proposed vindicates the notion that assigning intrinsic prudential value to more things makes a theory of welfare more complex.

Condition (c), meanwhile, excludes redundant claims by removing propositions that are entailed by others in the same theory. This ensures that nothing gets double-counted. It also forecloses a way in which every theory could be expressed as an infinite number of fundamental claims. For notice that if we begin with some proposition meeting conditions (a) and (b), such as:

$$(\forall x)(Px \supset Vx)$$

We can then infer another proposition meeting conditions (a) and (b), such as:

$$(\forall y)[(\forall x)(Px \supset Vx) \vee By]$$

From which we could then infer another proposition meeting conditions (a) and (b), such as:

$$(\forall z)\{(\forall y)[(\forall x)(Px \supset Vx) \vee By]\} \vee Cz\}$$

And so on *ad infinitum*. As we've seen, the first of these propositions (under a certain interpretation) is one of BPH's fundamental claims. Without condition (c), the choice of whether to count the latter two propositions as fundamental claims of BPH would be arbitrary, and so on for further propositions derived in the same way. With condition (c), we block this infinite regress before it begins.

I do not mean to suggest that mine is the single correct criterion of theoretical simplicity; only that it describes a property of sets of propositions that it is intuitive to regard as theoretical simplicity of a sort. Nothing substantive hangs on the use of the particular term "simplicity" (or related terms like "simple", "complex", or

“complicated”). Any reader who objects that “simplicity” in the context of philosophical theories (or theories generally) properly refers to something else may simply regard my use of the term as a bit of technical shorthand, replaceable by any other term they would prefer.

Why do I adopt a syntactic criterion of simplicity rather than an ontological one—particularly in light of Moen’s emphasis on the “ontological cost” of non-hedonist welfare theories? The main reason is that, while it is intuitive that such theories are more complex, it is unclear whether theories that introduce non-hedonic intrinsic values really do thereby impose an ontological cost. For illustration, compare prudential hedonism to a theory that assigns intrinsic positive prudential value to two things, pleasure and love. This theory shares with prudential hedonism a commitment to the existence of pleasure, but it also refers to something in the world—love, whatever sort of thing that is, ontologically speaking—that prudential hedonism does not. In that sense the pleasure-plus-love theory does have a larger and therefore less simple ontology. But prudential hedonists do not deny the existence of love; we just deny that it has intrinsic prudential value. So declining to refer to it in our theory of welfare doesn’t make for a simpler ontology overall. It does, however, make for a simpler theory of welfare in a syntactic sense, as the pleasure-plus-love theory contains a proposition (the one asserting the intrinsic positive prudential value of love) that prudential hedonism does not.

For the remainder of this chapter, all references to simplicity will be understood in terms of the syntactic criterion given above. And shortly I will attempt to show that

BPH is simpler than its competitors by this very standard. But first, we need some idea of how a theory of welfare being simpler could give us epistemic reason to accept it.

4.1.2 Epistemic Probability

Briefly, here is what I have in mind. For any two theories of welfare, where the epistemic probabilities of their fundamental claims are equal across the board (both within and between the theories being compared), the one that contains fewer fundamental claims will have a higher epistemic probability overall. So, under these conditions, we have epistemic reason to prefer simpler theories.³⁴

When I refer to the epistemic probability of a theory or claim, I am talking about the kind of probability at issue when we say that a claim is more or less plausible, or reasonable to accept, or likely to be true. Epistemic probability, in this sense, is a measure of how justified one would be in believing a proposition. As I understand it, epistemic probability is measured relative to one's evidence. This means that if two agents are contemplating accepting one and the same proposition, but each agent has different evidence at their disposal, the epistemic probability of that proposition may well be different for each of them. That said, in what follows, I will assume that we are working from the same body of evidence.³⁵

³⁴ Swinburne (1997) defends a view of this general sort about scientific theories.

³⁵ Climenhaga (2020: p. 3214) helpfully defines epistemic probability as follows, where A is some proposition and B is some evidence (or purported evidence) for A:

The epistemic probability of A given B—notated $P(A|B)$ —is a relation between the propositions B and A. It is the degree to which B supports A, or makes A plausible. Entailment is a limiting case of this relationship; if B entails A, then $P(A|B) = 1$: It constrains rational degrees of belief, in that, if $P(A|B) = n$; then someone with B as their evidence ought to be confident in A to degree n.

If you know that some proposition is a logical truth, or recognize a proposition as self-evident, then the epistemic probability of that proposition for you is one. If you know that some proposition is a logical contradiction, or you recognize some proposition as self-evidently false, then the epistemic probability of that proposition for you is zero. Every other proposition (or at least every proposition that you are in a position to understand) has an epistemic probability somewhere in between.

I am proposing that for any two theories of welfare, the one that contains fewer fundamental claims will have a higher epistemic probability in this sense. This proposal depends on at least two potentially contentious assumptions:

- (1) The fundamental (non-conjunctive, non-entailing) propositions of theories of welfare have in-principle quantifiable epistemic probabilities.
- (2) For any two fundamental propositions in a theory of welfare, A and B, the epistemic probability of (A&B) is lower than the epistemic probability of either A or B taken individually.³⁶

I have nothing approaching a decisive argument for either of these. But I will give some further explanation of each.

(1) says that the fundamental claims made in theories of welfare have epistemic probabilities in the sense just given. These would be claims like “all achievement has positive intrinsic prudential value” and “all pain has negative intrinsic prudential value”.

Achinstein (2001: pp. 96-113) is the rare philosopher who defends the notion of epistemic probability in an objective sense. In his view, objective epistemic probability—or objective reasonableness of belief—undergirds subjective epistemic probability. I see no conflict between this understanding of epistemic probability and what I argue in this chapter; I omit mention of it in the main text because it is, so far as I can tell, a fringe view (not to say an unreasonable one).

³⁶ By the definition I have given here, fundamental claims are non-entailing, meaning that no fundamental claim logically entails any other. So we need not be concerned with situations in which A is a subset of B or *vice versa*.

If you believe it makes sense to say that one of these propositions is more likely to be true than the other, or that they are about equally probable, and you mean this as something other than a way of expressing your levels of confidence in their truth, then you are already working with a concept of epistemic probability as I understand it. Now, consider the following claims:

- i. Your pleasures are intrinsically good for you.
- ii. Remembering your past pleasures is intrinsically good for you.
- iii. Someone else remembering their past pleasures is intrinsically good for you.

I regard (i) as more epistemically probable than (ii), and (ii) as more epistemically probable than (iii). I also think that the gap in epistemic probability between (ii) and (iii) is much larger than the gap in epistemic probability between (i) and (ii).

Here is why. I doubt that remembering past pleasures is intrinsically good for me. But I do believe that at least some of my own private mental states can be intrinsically good for me, namely my pleasures. My memories of my pleasures are also private mental states of mine, their contents concern states I regard as intrinsically valuable for me, and they are often pleasant to have. So while I do not think they are good for me in themselves, they do at least have some prominent similarities with, and close relations to, states that I do think are good in this way. Another person's memories of their pleasures, by contrast, are inaccessible to me, do not necessarily have anything to do with me, and, not being mine, cannot be pleasant for me to have. They therefore do not seem in any way like the sorts of things that could be intrinsically good for me.

For these reasons, I believe that there is a gap in epistemic probability between (i) and (ii), and another, larger gap in epistemic probability between (ii) and (iii). If this is right, then in at least some cases, disparities in epistemic probability come in sizes that cannot be captured on a merely ordinal scale. Some kind of in-principle, non-ordinal quantifiability is called for. This would not show that all fundamental welfare-theoretic propositions have in-principle quantifiable epistemic probabilities, but is a start, and will have to suffice for now.

Assumption (2) suggests that adding fundamental claims to a theory of welfare lowers its total epistemic probability, while removing them raises it. This is how other sets of probabilistic claims behave in cases where all are above zero and below one (just think of multiple flips of a coin or rolls of a die). The fundamental claims of welfare theories are not self-evident or logically true, nor are they internally contradictory, so they do indeed all have epistemic probabilities above zero and below one. What (2) proposes, then, is just that the epistemic probabilities of these claims interact with one another as probabilities standardly do.

I have proposed that for any two theories of welfare, where the epistemic probabilities of their fundamental claims are equal across the board (both within and between the theories being compared), the one that contains fewer fundamental claims will have a higher epistemic probability overall. Now, if epistemic probabilities interact with one another as probabilities standardly do, then this will not be the only situation in which the simpler of two welfare theories will have the higher epistemic probability. For instance, in any case in which the epistemic probability values of the fundamental claims

of one theory form a proper subset of the epistemic probability values of the fundamental claims of another theory, the former, simpler theory will have the higher epistemic probability overall.³⁷³⁸

So why do I focus on this one sort of case, in which the fundamental claims of the theories being compared have epistemic probabilities that are equal across the board? Because I believe that this a reasonable assessment of epistemic situation before us in this chapter. Indeed, the goal of this dissertation so far has in effect been to create this situation

³⁷ Here's a toy example. Suppose we are comparing two theories, T_1 and T_2 . T_1 contains exactly three fundamental claims, with epistemic probability values of 0.1, 0.2, and 0.3. T_2 contains exactly four fundamental claims, with epistemic probability values of 0.1, 0.2, 0.3, and n . The epistemic probability values of the fundamental claims of T_1 form a proper subset of the epistemic probability values of the fundamental claims of T_2 : $\{0.1, 0.2, 0.3\} \subset \{0.1, 0.2, 0.3, n\}$. In this case, as long as $n < 1$, the total epistemic probability of T_1 will be greater than that of T_2 .

³⁸ What about cases in which one theory's fundamental claims are a proper subset of another theory's fundamental claims? We might think that this is a relevant sort of case, since BPH's fundamental claims may appear to be a proper subset of the fundamental claims contained in any objective list theory that assigns positive intrinsic prudential value to pleasure, negative intrinsic prudential value to pain, and positive or negative intrinsic prudential value to least one other thing besides. For example, we might think that BPH's fundamental claims are, or at least could be, a proper subset of the fundamental claims contained in an objective list theory that assigned negative intrinsic prudential value to all and only pain, and positive intrinsic prudential value to all and only pleasure and autonomy.

But this would be incorrect. Again, let P stand for the predicate "is a pleasure", A stand for the predicate "is an instance of autonomy", and V stand for the predicate "has positive intrinsic prudential value", and consider just fundamental claims about positive intrinsic prudential value. BPH's fundamental claims in this area are these:

$$(\forall x)(Px \supset Vx), (\forall x)(Vx \supset Px)$$

Whereas the pleasure-plus-autonomy theory's fundamental claims in this area are:

$$(\forall x)(Px \supset Vx), (\forall x)(Ax \supset Vx), (\forall x)[Vx \supset (Px \vee Ax)]$$

As we can see, BPH contains a fundamental claim that the pleasure-plus-autonomy theory lacks: $(\forall x)(Vx \supset Px)$, i.e. the claim that only pleasure has intrinsic prudential value. The latter theory's version of this claim is $(\forall x)[Vx \supset (Px \vee Ax)]$, i.e. the claim that only pleasure and autonomy have intrinsic prudential value, which does not entail $(\forall x)(Vx \supset Px)$. So BPH's fundamental claims are not a proper subset of the pleasure-plus autonomy theory's fundamental claims.

for the reader. To see how, let's retrace the dialectic so far with the concept of epistemic probability in mind.

I began in chapter 1 with some remarks on what is fundamentally appealing about prudential hedonism, desire theory, and objective list theory as approaches in the philosophy of welfare. I proposed, with reference to specific theories belonging to these categories, that each approach is fundamentally appealing in that it can offer a recognizable picture of a life that is good for its subject under normal material and psychological circumstances. Initially, then, there appear to be several theories of welfare making sharply distinct claims about what has intrinsic prudential value that nonetheless have equal *prima facie* credibility. So, to be justified in favoring or disfavoring any of these theories relative to others, we need arguments.

In chapters 2 and 3, I discussed several arguments that purport to give us reason to reject BPH. If successful, these arguments would (in effect) lower the epistemic probability of BPH relative to its rivals. For example, the experience machine argument, if successful, would lower the total epistemic probability of BPH by lowering the epistemic probability of its claim that only pleasure has positive intrinsic prudential value (or the claim that only pain has negative intrinsic prudential value, or both). But I believe that I showed these arguments to be unsuccessful. If I'm right about this, then BPH's epistemic probability has emerged unscathed. This would leave us, the outset of this chapter, in the same epistemic position as we were at the end of chapter 1: having no particular reason to favor BPH over any of its main competitors or *vice versa*. A reasonable way of assessing this situation, I think, is as one in which the fundamental claims of these

theories have equal epistemic probabilities across the board.³⁹ (Note, however, that the claim of exact equality is not strictly necessary. All of what follows works just as well so long as the epistemic probabilities of BPH's fundamental claims are not lower than those of its rivals.)

In summary: if we accept (1) and (2), and if I have been successful in this dissertation so far, then BPH being simpler than its rivals would give us an epistemic reason to accept it. This brings us to the challenge of showing that BPH is indeed the simpler view.

4.2 The Simplicity of Basic Prudential Hedonism

In this section I will make a *prima facie* case for the claim that BPH is simpler than rival objective list theories, desire theories, and alternative forms of prudential hedonism. I do not claim to be able to make a precise count of the fundamental claims in any theory of welfare. But this is not necessary. It will suffice to show that, however many fundamental claims BPH and its rivals make, each of those rivals must contain more such claims than BPH does.

4.2.1 Minimal Viability Conditions for Theories of Welfare

Before moving on, I should specify which theories I consider to be BPH's rivals (for the purpose of simplicity comparisons). I have in mind theories of welfare that meet the following four "minimal viability conditions":

³⁹ Granted, I cannot rule out that the reader is aware of a plausible anti-BPH argument that I have not rebutted, either directly or implicitly by way of casting doubt on assumptions of qualitative or prudential insight. For such a reader, the epistemic probability of BPH may be low relative to one or more of its rivals. All I can say is that I do not know of any such arguments myself.

(Consistency) The theory is internally logically consistent.

(Axiology) The theory contains a set of fundamental claims assigning positive intrinsic prudential value to one or more things, and another set of fundamental claim assigning negative intrinsic prudential value to one or more things.

(Description) The theory provides some non-evaluative account of each type of thing it claims has intrinsic prudential value.

(Comparison) The theory provides some in-principle account of how to determine, for any two lives, which (if any) is prudentially better for its subject.⁴⁰

The consistency condition removes theories of welfare that are internally inconsistent from the comparison class, such as unrestricted actualist desire theories (see §1.2.2). Hence I will only be comparing BPH to restricted versions of desire theory.

⁴⁰ It is rare for philosophers of welfare to explicitly state a set of conditions that they think a theory needs to meet to be plausible. The one exception I know of is Sumner (1996: pp.12-18), who says that a successful theory of well-being must be descriptively adequate, which in turn requires meeting at least four criteria. What follows are my brief summaries of these criteria, which I hope do justice to Sumner's longer explanations:

Fidelity: The theory should be faithful to our pretheoretical beliefs about well-being, adjusted for the centrality of those beliefs to our network of preanalytic convictions.

Generality: The theory should provide truth conditions both for claims about a person's welfare level at a given time and for claims about changes in welfare, positive or negative. It should be able to do this for all kinds of people and for some non-human creatures, and for groups as well as individuals.

Formality: A theory of well-being must not confuse the nature of well-being (unitary) with its sources (plural). It should be able to tell us, for any source of well-being, what makes it a source of well-being.

Neutrality: The theory should not have built into it a bias toward any particular goods or way of life.

These conditions are more stringent than the ones I propose in this chapter, which do not even require that theories assign value in a manner faithful to pre-theoretic beliefs about welfare.

The axiology condition rules out as irrelevant theories that assign positive but not negative intrinsic prudential value, or *vice versa*. This condition is justified by the fact that theories violating this constraint are implausible regardless of their specific content.

To see why, consider a theory that claims that all and only the satisfaction of actual, rational desires are intrinsically good, but does not say that the frustration of these desires, or anything else, is intrinsically bad. This theory might well be simpler than BPH. But it has the implausible implication that it is not possible for a person's total welfare to become lower than it was at any previous point in time. On this view, we would rack up more and more welfare with each instance of rational desire-satisfaction, and no injury, failure, or humiliation could budge the tally in the opposite direction. Such things would make our lives worse for us only insofar as they caused us to accumulate rational desire-satisfactions more slowly. It would follow that no matter how strong your rational desire to avoid pain, a year of horrific torture could not have a more deleterious effect on the quality of your life for you than, say, a year spent in a desire-free coma. (If this does not strike the reader as implausible, then I do not know what to say.)

Any theory of welfare that assigns positive intrinsic prudential value to something while denying that anything has negative prudential value in itself will face this sort of objection. And any theory of welfare that assigns negative intrinsic prudential value to something while denying that anything has positive prudential value in itself will face a similar sort of objection, as it will imply that it is not possible for a person's life to get better for them than it was at any previous time, except perhaps by slowing the rate at which it gets worse. I conclude that for a theory of welfare to even minimally viable, it

must say that some things are good for us in themselves and that some things are bad for us in themselves.

A theory must meet the description and comparison conditions, meanwhile, in order for it to present a recognizable picture of a life that is good for its subject under normal material and psychological circumstances.

Without some non-evaluative account of the things to which the theory assigns intrinsic prudential value, it will not be at all clear what the theory is asserting. For example, if a theory claims that pleasure is intrinsically good, but takes no stance whatsoever on the non-evaluative nature of pleasure, then it will not be clear what the content of that theory is (if indeed it has any). Thus the description condition obliges prudential hedonists to say something about what pleasure and pain are, informed desire theorists to say something about what desire is and what it takes for a desire to be suitably informed, objective list theorists to give some non-evaluative description of each item on their lists, and so on.

Moreover, without some account of how to determine which of a pair of lives is prudentially better for its subject, even a theory that meets the axiology and description conditions in a more or less reasonable way is not yet sure to be even remotely plausible.⁴¹ Consider a version of prudential hedonism that meets the former two conditions just as

⁴¹ I am not saying that for a theory to be minimally viable, it must issue a definite verdict about every suitably described life comparison (though BPH does do this). Nor am I saying that for a theory to be minimally viable it must reject the possibility that there will be cases in which there is no fact of the matter as to which of two lives is prudentially superior. I am saying that to be minimally viable, a theory cannot be silent on the matter; it must make one or more fundamental claims about which lives are better than which.

BPH does, i.e. with the evaluative and phenomenological theses. This theory would tell us that pleasure and pain are each distinct phenomenal properties, and that they have positive and negative intrinsic prudential value (respectively). But it would not tell us how these values combine to determine the quality of a person's life for them. There are more or less plausible ways of answering this question, so until we have an answer to it, we cannot tell whether the theory paints a recognizable picture of a life that is good for its subject under normal material and psychological circumstances.⁴²

These conditions set a low bar for theories of welfare. They do not demand that the actual content of a theory be at all reasonable. But meeting them ensures that a theory is internally consistent, and is necessary for a theory to be initially plausible. The conditions also set a floor for how simple a minimally viable theory of welfare can be. However few fundamental claims a theory may contain, it must have enough to meet the axiology, description, and calculation theses. So, in the remainder of this section, I will be assessing the simplicity of BPH only against minimally viable objective list theories, desire theories, and alternative forms of prudential hedonism.

4.2.2 Simplicity Comparisons

I begin with objective list theories. These theories assign intrinsic, objective prudential value to multiple things. But, all else being equal, a theory of welfare is simpler to the extent that it assigns intrinsic prudential value to fewer things. BPH assigns such

⁴² Here is an example of an implausible way of answering this question. A hedonist theory of welfare could say that the intrinsic prudential value is extremely slight compared to that of pleasure, such that the negative value of a year of horrific torture would be outweighed by the positive value of a brief mildly pleasant sensation.

value to only two things: pleasure on the positive side and pain on the negative. In every case that I know of, objective list theorists assign intrinsic prudential value to more than two things.⁴³ These theories therefore contain more fundamental claims than BPH does.

I do not know how to translate every one of a theory's fundamental claims into first-order logic in order to make a precise count. But even my very modest formal logic skills suffice for demonstrating the above point. Consider a toy example. BPH is simpler than a theory that assigns intrinsic positive prudential value to both pleasure and autonomy. This is because assigning value to autonomy in addition to pleasure requires making at least one additional fundamental claim. Recall that BPH's fundamental claims about positive intrinsic prudential value are just these two, where P is the predicate for pleasure and V the predicate for having intrinsic positive prudential value:

$$(\forall x)(Px \supset Vx), (\forall x)(Vx \supset Px)$$

The first of these claims says that all pleasures have intrinsic positive prudential value, while the second says that only pleasures have such value.

Meanwhile, the pleasure-plus-autonomy theory must invoke an additional predicate for autonomy (A), making for at least these three fundamental claims about positive intrinsic prudential value:

$$(\forall x)(Px \supset Vx), (\forall x)(Ax \supset Vx), (\forall x)[Vx \supset (Px \vee Ax)]$$

⁴³ For example, Hurka (2011) says that pleasure, knowledge, achievement, virtue, and friendship are all intrinsically good for us; Fletcher (2013) proposes that the list of objective intrinsic prudential goods consists of pleasure, friendship, achievement, happiness, self-respect, and virtue; Hooker (2015) argues that the list consists of pleasure, friendship, significant achievement, important knowledge, and autonomy.

The first two claims here assign positive intrinsic prudential value to all pleasures and all instances of autonomy, while the third says that nothing other than pleasure and autonomy has such value. We can extrapolate from this example to any other objective list theory.

I turn next to desire theory. As I noted earlier, a logically consistent version of desire theory must restrict the range of desires that can affect a person's welfare. Such restrictions require the theory to make additional fundamental claims, in order to specify what properties the right sort of desire-objects will have or lack. For example, compare unrestricted actualist desire theory (setting its internal inconsistency aside for the moment) with a view on which only the satisfactions of actual, rational desires are good for us in themselves. However many fundamental claims the unrestricted desire theory contains, this restricted view will have to have at least one more, in order to restrict the set of intrinsically good desire-satisfactions to only the rational ones.

Here is an illustration. Let D stand for the predicate "is an instance of actual desire-satisfaction", R stand for the predicate "is rational", and V once again stand for "has positive intrinsic prudential value". On this interpretation, unrestricted actualist desire theory makes these two fundamental claims about what is good for us in itself:

$$(\forall x)(Dx \supset Vx), (\forall x)(Vx \supset Dx)$$

The restricted view, however, makes these three fundamental claims about what is good for us in itself:

$$(\forall x)[(Dx \& Rx) \supset Vx], (\forall x)(Vx \supset Dx), (\forall x)(Vx \supset Rx)^{44}$$

Once again, we can extrapolate to other views of the same sort. So, while unrestricted actualist desire theory might be just as simple as BPH, all logically consistent versions of the view will be less so.

The same basic points apply to ideal desire theories. These theories say that only the satisfactions and frustrations of desires the agent would have under ideal conditions, such as conditions of full information and rationality, have intrinsic prudential value. Such views must make additional claims ascribing the right properties to the ideal conditions.

Finally, we should consider some alternative forms of prudential hedonism. Having already argued against attitudinal theories of hedonic states in chapter 2, I will not attempt to show that BPH is simpler than attitudinal versions of prudential hedonism. I will focus my attention on two sorts of deviation from BPH: restricting the range of intrinsically valuable hedonic states, and positing discontinuities in the hedonic calculus. Each of these deviations invariably introduces complexity.

First, the restricted forms of prudential hedonism. Consider, for example, a hedonist view on which only morally innocent pleasures have positive intrinsic prudential value. The comparison between this view and BPH is analogous to the comparison between restricted and unrestricted versions of desire theory discussed above. The morally restricted version of prudential hedonism will have to contain one

⁴⁴ Note that the proposition $(\forall x)[Vx \supset (Dx \& Rx)]$ entails the conjunction $[(\forall x)(Vx \supset Dx) \& (\forall x)(Vx \supset Rx)]$, and therefore cannot be a fundamental claim by the definition given in §4.1.1.

additional fundamental claim specifying that only pleasures with a certain property (moral innocence) are intrinsically good for us.⁴⁵ We can extrapolate from this to any other theory that assigns intrinsic prudential value to only a subset of hedonic states. All such views will be less simple than BPH.

Second, versions of prudential hedonism that posit discontinuities in the hedonic calculus. By now we are familiar with one such view: qualitative hedonism in the vein of Mill (2001) and Crisp (2006). Qualitative hedonism has it that there are quality distinctions among pleasures; Crisp, for example, holds that no amount of lower-quality pleasure can equal or exceed any amount of higher-quality pleasure in hedonic magnitude. Capturing this will require additional fundamental claims, at least one for each discontinuous level of hedonic quality.

The same will be true for any other theory that posits discontinuity in the hedonic calculus, whether it is a version of prudential hedonism or not. For example, consider the “disunified phenomenological thesis” (DPT) discussed at §2.2.2. The DPT says that what makes a mental state a pleasure or a pain is the way that it feels (and not a relation to something else, like an attitude), but not all pleasures share a single phenomenal property, and neither do all pains. Instead, pleasure and pain are unified as categories by phenomenological similarities (perhaps on the model of Wittgensteinian family

⁴⁵ Let P stand for the predicate “is a pleasure”, M stand for the predicate “is morally innocent”, and V stand for the predicate “has positive intrinsic prudential value”. Prudential hedonism that restricts the range of intrinsically valuable pleasures to only the morally innocent ones will have to make the following three claims about what is good for us in itself, whereas, as we have seen, BPH makes only two in this regard:

$$(\forall x)[(Px \& Mx) \supset Vx], (\forall x)(Vx \supset Px), (\forall x)(Vx \supset Mx)$$

resemblances). This view implies that the hedonic calculus is discontinuous. If it were continuous, then there would have to be a single property shared among all pleasures that makes them count as pleasures (and another among all pains that makes them count as pains), which is precisely what the DPT denies. Hence the DPT is less simple than the phenomenological thesis, and any form of prudential hedonism built on the DPT will be less simple than BPH.

The upshot is that BPH is simpler than minimally viable objective list theories, restricted forms of desire theory, and versions of prudential hedonism that restricts the range of intrinsically valuable hedonic states or posits discontinuity in the hedonic calculus. From this and the assumptions discussed in the previous section, it follows that BPH has a higher epistemic probability than these views. This gives us an epistemic reason to favor BPH over these competitors. I turn now to three challenges to this argument.

4.3 Challenges

4.3.1 The Problem of Selecting Predicates

To determine whether one theory contains fewer fundamental claims than another, we first need to settle on a common set of predicates with which to express these theories. In other words, we need a common language of fundamental properties and relations. But different languages can produce different results: T_1 may come out simpler than T_2 with one set of predicates but not another. If there is no right way of choosing what language to use for the sake of measuring relative simplicity, then there is no fact of the matter as to which of two theories is really simpler.

Now consider the following scenario. Someone proposes an objective list theory of welfare on which only pleasure and autonomy have intrinsic positive prudential value, and they coin a new term, “plautonomy”, defined such that anything that is an instance of either pleasure or autonomy is an instance of plautonomy. He then assigns this term its own predicate (L). With this predicate in hand, the pleasure-plus-autonomy theory need not make any more fundamental claims than BPH. Its claims about positive intrinsic prudential value can be expressed as follows:

$$(\forall x)(Lx \supset Vx), (\forall x)(Vx \supset Lx)$$

With the plautonomy predicate in our language, BPH and the pleasure-plus-autonomy view make an equal number of fundamental claims about what has positive intrinsic prudential value. If this hypothetical philosopher proposes that only pain has negative intrinsic prudential value, the resulting theory would appear to be no more complex than BPH overall. So it seems that if I want to say that BPH is simpler than the pleasure-plus-autonomy view, I need some principled basis for selecting a language without the plautonomy predicate.

One possibility would be to demand that the language chosen for the sake of a simplicity comparison contain only as many predicates as needed to express the two theories being compared, and no more. But this doesn’t solve the problem. In some cases, multiple languages will meet this demand, and still issue different verdicts about comparative simplicity. For example, we could express BPH’s claims about positive intrinsic prudential value without a predicate for pleasure, using only the predicates for autonomy (A) and plautonomy (L):

$$(\forall x)[(Lx \ \& \ \sim Ax) \supset Vx], (\forall x)(Vx \supset Lx), (\forall x)(Vx \supset \sim Ax)$$

In this language BPH could turn out to be less simple than the pleasure-and-autonomy theory.

The only option seems to be to demand that the predicates of the language we use for simplicity comparisons correspond to the real fundamental properties and relations out there in the world—so-called “natural kinds”. But then, of course, we have the problem of how to figure out what these properties and relations are.

Opponents could leverage this problem against BPH in one of at least three ways. First, they could claim that a proper account of the real fundamental properties and relations out there in the world would show that their preferred theory of welfare is just as simple as BPH (or even simpler). For instance, one could argue that it is the satisfactions and frustrations of rational desires that have intrinsic prudential value (positive and negative, respectively), and that each of these is a natural kind for which a single predicate would be appropriate. Second, they could argue that there are no natural kinds, so there is no fact of the matter as to whether one theory is simpler than another, at least according to the criterion I have proposed. Third, they could argue that we rationally ought to suspend judgment about the existence of natural kinds, and therefore we ought to suspend judgment about whether BPH is simpler than its rivals.

I do not know how to determine whether there are natural kinds and what they might be, so I do not know how to answer any of these objections. I will note, however, that even if successful, such an objection would not give us reason to reject BPH. It would

only refute my case for accepting it over its major competitors, returning us to the stalemate with which this chapter began.

4.3.2 Universal Quantifiers and Skepticism

The notion of epistemic probability, as applied to theories (philosophical or otherwise), faces a troubling general objection. Any theory on any subject will include some universal claims. As we have seen, BPH contains the claim that all pleasures have intrinsic prudential value:

$$(\forall x)(Px \supset Vx)$$

But notice that a universal claim like this one is logically equivalent to the conjunction of all its substitution instances:

$$[(Pa \supset Va) \ \& \ (Pb \supset Vb) \ \& \ (Pc \supset Vc) \ \dots]$$

And in principle, the number of substitution instances is infinite, since any object at all could be substituted for the variable “x”. The objection, then, is that all theories with universal claims in them—which is to say all theories—are in fact equivalent to an infinitely long set of propositions, and therefore all theories containing non-analytic universal propositions have epistemic probabilities that asymptotically approach zero.

This is not an objection that applies to theories of welfare in particular, let alone BPH specifically. It is a general objection to non-ordinal measures of epistemic probability. It points out that if the substitution instances of universal claims can have epistemic probabilities above zero and below one, then any universal claim that is neither a logical truth nor a contradiction will be vanishingly improbable. This presents a choice: give up on the idea of non-ordinal epistemic probability—that is, reject the idea that there is any

fact of the matter as to how much more justified one proposition is than another—or embrace the idea that when you assent to a non-formal theory, be it scientific or philosophical, you are assenting to a set of claims that is scarcely epistemically preferable to a necessary falsehood. I find both of these difficult to believe, which suggests to me that the objection itself is has gone wrong somewhere. But, I confess, I do not know how it is mistaken. I leave it to opponents of BPH to decide whether this challenge is one whose implications they themselves are prepared to accept.

4.3.3 Is BPH Really Simpler?

One might agree that BPH is the simplest minimally viable theory of welfare, on precisely the grounds I have suggested here, and yet doubt that it is the minimally viable theory of welfare that makes for the smallest set of welfare-related theoretical commitments. For it may be that defending BPH against the objections discussed in previous chapters requires so many additional claims that adopting the view comes with no advantage in simplicity overall.⁴⁶

Recall that in chapter 3, I contrasted two hypotheses about our intrinsic prudential intuitions: the strong insight hypothesis, which says that our intrinsic prudential intuitions reliably track the truth about what really has or lacks intrinsic prudential value, and the pro-hedonist hypothesis, which says this:

⁴⁶ Thanks to Kobi Finestone for calling this objection to my attention.

The Pro-Hedonist Hypothesis: BPH is true, and we tend to correctly assign positive intrinsic prudential value to pleasure. But we also tend to mistakenly assign positive intrinsic prudential value to other things, just in case they meet these three conditions:

- (c) We observe them to be consistently conducive to net pleasure for people in general.
- (d) They are not categorically instrumental.
- (d) They are not activities that are usually done for some external end.

Our intrinsic prudential intuitions are the products of tacit inferences from features of the given scenarios plus these prior assignments of intrinsic value. (*Mutatis mutandis* for pain and negative intrinsic prudential value.)

The pro-hedonist hypothesis looks a lot more complicated than the strong insight hypothesis. The former is much more succinct: it just says that we have the intrinsic prudential intuitions that we do because we have some capacity for generating intrinsic prudential intuitions with reliably true contents. On the other hand, the pro-hedonist hypothesis goes on at comparatively great length, specifying the various conditions under which we will have intuitions of a pro- or anti-hedonist sort. If defending BPH against evaluative objections requires taking on the pro-hedonist hypothesis, or something like it, while opponents can stick with the claim of strong prudential insight, then BPH's claim to greater simplicity would seem to evaporate.

This appearance would be misleading. Simplicity is only a *ceteris paribus* theoretical virtue: it becomes relevant only when all else is equal between the sets of claims being assessed. But not all is equal between the strong insight hypothesis and the pro-hedonist hypothesis, because the latter is more informative: it offers an account of both accurate intrinsic prudential intuitions and erroneous ones. The strong insight hypothesis only tells about the former.

Now, if the strong insight hypothesis were revised to say that our intuitions track the truth about intrinsic prudential value without fail, then it would just as informative as, and simpler than, the pro-hedonist hypothesis. But this would be to say that all intrinsic prudential intuitions have true contents, which would imply that different people's intrinsic prudential intuitions never contradict one another. This is not what we observe. Defectors from the intuitive consensus on cases like the experience machine may be rare, but they do exist. We therefore cannot assess the relative simplicity the total welfare-theoretic worldviews of basic prudential hedonists and their opponents until the strong insight hypothesis is bolstered by some account of cases in which people's intrinsic prudential intuitions miss the mark.⁴⁷

I must add that though this objection is mistaken, it shows something important. I have tried to shield my defense of BPH from empirical disconfirmation as much as possible. This objection reveals a limit of this effort. Whether accepting and defending BPH really makes for a simpler welfare-related belief set depends in part on how simple the pro-hedonist hypothesis is relative to the equally plausible and informative hypotheses about our intrinsic prudential intuitions that are available to proponents of rival views. This, in turn, depends on what data these hypotheses need to account for, i.e. what intrinsic prudential intuitions people actually have, which is an empirical matter. I suspect that anti-hedonist accounts of our prudential intuitions need not be any simpler

⁴⁷ This approach would also encounter difficulties with intuitions that tell against non-hedonist theories of welfare in general. For example, see Pummer (2017).

than pro-hedonist ones, all else being equal; but in the absence of data, I can only speculate.

4.4 Conclusion

If the approach I have outlined in this chapter is right, then basic prudential hedonism is simpler than its competitors in a way that makes it more likely to be true. However, as we have seen, the argument depends on a number of questionable assumptions. It is open to opponents of BPH to exploit this. But doing so would come at the cost of taking on theoretical commitments far afield from the topic of well-being, concerning matters such as epistemic probability and the existence of natural kinds. Refuting the argument in this chapter, then, may require more philosophical breadth than is typical of arguments in the philosophy of welfare.

5. Conclusion

The aim of this dissertation has been twofold: to show that basic prudential hedonism (BPH) is at least as defensible as other positive theories of welfare, and to show that it is in fact preferable to its competitors, such that it is the positive theory of welfare that we rationally ought to accept.

In chapter 1, I showed that BPH was at least the equal of major desire satisfaction and objective list theories in its fundamental appeal. In chapter 2, I gave a positive argument for the phenomenological thesis, and defended BPH against a number of descriptive (non-evaluative) objections. There, I also gave a case for remaining neutral on the choice between separate experience and hedonic tone theories of hedonic states. Then, in chapter 3, I defended BPH against a number of objections based on one or more evaluative premises. These chapters represent my effort to show that BPH is at least as defensible as its competitors among positive theories of welfare. Of course, I do not claim to have addressed every extant descriptive or evaluative objection to BPH. But I believe that the approaches I outline in chapters 2 and 3, particularly concerning the assumptions of quantitative, qualitative, and prudential insight, should be applicable to many objections not explicitly addressed in these pages.

My effort to meet the second, more ambitious aim—showing that BPH is the theory of welfare we rationally ought to accept—was confined to chapter 4. There, I outlined an argument to the effect that we rationally ought to accept BPH because, of all the minimally viable theories of welfare, it is the simplest (in a specific syntactic sense), and therefore the most likely to be true (in a specific epistemic sense).

But, as we have seen, the success of the argument outlined in chapter 4 is contingent on some contentious empirical, metaphysical, and epistemological assumptions, defense of which is beyond the scope of this work. Whether BPH really makes for the simplest worldview overall depends, in part, on what intuitions people really have about what has intrinsic prudential value. And whether my overarching approach of arguing for BPH on the grounds of its simplicity succeeds depends on controversial assumptions in fundamental philosophy.

I see no way to avoid this. But there is a silver lining, which is that BPH's opponents are in the same position. Far from being a settled matter of philosophical common sense, the case against basic prudential hedonism relies on a number of questionable assumptions—about hedonic insight, prudential insight, what intuitions we generally have, and perhaps ultimately about the fundamental logical constituents of theories of welfare. If there is one accomplishment I can claim for this dissertation, it is that it sheds some light on this fact.

References

- Achinstein, P. (2001). *The Book of Evidence*. Oxford: Oxford University Press.
- Alston, W.P. (1967). "Pleasure". In Paul Edwards (ed.), *The Encyclopedia of Philosophy*, Vol 6. New York: Macmillan.
- Baber, H.E. (2008). "The Experience Machine Deconstructed". *Philosophy in the Contemporary World*, 15, 133-138.
- Barber, A. (2011). "Hedonism and the Experience Machine". *Philosophical Papers*, 40, 257–278.
- Bentham, J. (1988). *An Introduction to the Principles of Morals and Legislation*. Amherst, NY: Prometheus Books.
- Bradley, B. (2007). "A Paradox for Some Theories of Welfare". *Philosophical Studies*, 133, 45-53.
- Bramble, B. (2013). "The Distinctive Feeling Theory of Pleasure." *Philosophical Studies*, 137, 201-217.
- Bramble, B. (2016). "A New Defense of Hedonism about Well-Being". *Ergo*, 3.
- Brandt, R. (1979). *A Theory of the Good and the Right*. Amherst, New York: Prometheus Books.
- Climenhaga, N. (2018). "Intuitions Are Used as Evidence in Philosophy". *Mind*, 127, 69-104.
- Climenhaga, N. (2020). "The Structure of Epistemic Probabilities". *Philosophical Studies*, 177, pp. 3213–3242.
- Brax, D. (2009). *Hedonism as the Explanation of Value*. Doctoral dissertation, University of Lund.
- Brentano, F. (2009). *The Origin of Our Knowledge of Right and Wrong*. Ed. Oskar Kraus, English edn. ed. Roderick Chisholm, trans. Roderick Chisholm and Elizabeth Schneewind. London: Routledge.
- Crisp, R. (2006). *Reasons and the Good*. Oxford: Oxford University Press.
- Darwall, S.L. (1983). *Impartial Reason*. Ithaca, New York: Cornell University Press.

- Dorsey, D. (2011). "The Hedonist's Dilemma". *Journal of Moral Philosophy*, 8, 173-196.
- Feldman, F. (1999). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge: Cambridge University Press.
- Feldman, F. (2006). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford: Oxford University Press.
- Feldman, F. (2012). "What We Learn from the Experience Machine". In R.M. Bader and J. Meadowcroft (Eds.), *The Cambridge Companion to Robert Nozick's Anarchy, State, and Utopia* (59-88). Cambridge: Cambridge University Press.]
- Fletcher, G. (2013). "A Fresh Start for the Objective-List Theory of Well-Being". *Utilitas*, 25, 206-220.
- Fletcher, G. (2016). *The Philosophy of Well-Being: An Introduction*. New York: Routledge.
- Gert, B. (2005). *Morality: Its Nature and Justification*. Oxford: Oxford University Press.
- Griffin, J. (1986). *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Oxford University Press.
- Goldstein, I. (1985). "Hedonic Pluralism". *Philosophical Studies*, 48, 49-55.
- Goodman, N. (1958). "The Test of Simplicity". *Science, New Series* 128, pp. 1064-1069.
- Hawkins, J. (2016). "The Experience Machine and the Experience Requirement". In G. Fletcher (Ed.), *The Routledge Handbook of the Philosophy of Well-Being*, 355-365. New York: Routledge.
- Haybron, D. (2008). *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.
- Heathwood, C. (2006). "Desire Satisfactionism and Hedonism". *Philosophical Studies*, 128, 539-563.
- Heathwood, C. (2007). "The Reduction of Sensory Pleasure to Desire". *Philosophical Studies*, 133, 23-44.
- Heathwood, C. (2019). "Which Desires Are Relevant to Well-Being?" *Nous*, 53, 664-688.

- Hewitt, S. (2010). "What Do Our Intuitions About the Experience Machine Really Tell Us About Hedonism?" *Philosophical Studies*, 151, 331-349.
- Hooker, B. (2015). "The Elements of Well-Being". *Journal of Practical Ethics*, 3, 15-35.
- Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, Maryland: Rowan & Littlefield.
- Huemer, M. (2007). "Compassionate Phenomenal Conservatism". *Philosophy and Phenomenological Research*, 74, 30-55.
- Hurka, T. (1993). *Perfectionism*. Oxford: Clarendon Press.
- Hurka, T. (2011). *The Best Things in Life: A Guide to What Really Matters*. Oxford: Oxford University Press.
- Kagan, S. (1992). "The Limits of Well-Being". *Social Philosophy & Policy*, 9, 169-189.
- Kagan, S. (1993). "Me and My Life". *Proceedings of the Aristotelian Society*, 94, 309-324.
- Katz, L.D. (1986). *Hedonism as Metaphysics of Mind and Value*. Doctoral dissertation, Princeton University.
- Kawall, J. (1999). "The Experience Machine and Mental State Theories of Well-Being". *Journal of Value Inquiry*, 33, 381-387.
- Kenny, A. (2003). *Action, Emotion, and Will*. London: Routledge.
- Klocksiam, J. (2009). *On the Measurability of Pleasure and Pain*. Doctoral dissertation, University of Massachusetts Amherst.
- Klocksiam, J. (2010). "Pleasure, Desire, and Oppositeness". *Journal of Ethics and Social Philosophy*, 2, 1-7.
- Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Kraut, R. (2007). *What is Good and Why*. Cambridge, MA: Harvard University Press.
- Lafleur, L. (1955). "In Defense of Ethical Hedonism". *Philosophy and Phenomenological Research*, 16, 547-550.
- Lin, E. (2016). "How to Use the Experience Machine". *Utilitas*, 28, 314-332.

- Lin, E. (2020). "Attitudinal and Phenomenological Theories of Pleasure". *Philosophy and Phenomenological Research*, 100, 510-524.
- Lukas, M. (2010). "Desire Satisfactionism and the Problem of Irrelevant Desires". *Journal of Ethics and Social Philosophy*, 4.
- Mayerfeld, J. (1992). *Suffering and Moral Responsibility*. Oxford: Oxford University Press.
- Mendola, J. (1990). "Objective Value and Subjective States". *Philosophy and Phenomenological Research*, 50, 695-713.
- Mill, J.S. (2001). *Utilitarianism*. Indianapolis: Hackett.
- Moen, O.M. (2013). "The Unity and Commensurability of Pleasure and Pain". *Philosophia*, 41, 527-543.
- Moen, O.M. (2016). "An Argument for Hedonism". *Journal of Value Inquiry*, 50, pp. 267–281.
- Murphy, M. (1999). "The Simple Desire-Fulfillment Theory". *Nous*, 33, 247-272.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Pummer, T. (2017). "Lopsided Lives". In Mark Timmons (ed.), *Oxford Studies in Normative Ethics*, pp. 275-296. Oxford: Oxford University Press.
- Rachels, S. (2004). "Six Theses About Pleasure". *Philosophical Perspectives*, 18, 247-267.
- Railton, P. (1986a). "Moral Realism". *The Philosophical Review*, 95, 163-207.
- Railton, P. (1986b). "Facts and Values". *Philosophical Topics*, 14, 5-31.
- Railton, P. (1989). "Naturalism and Prescriptivity". *Social Philosophy and Policy*, 7, 151-174.
- Rosati, C.S. (1996). "Internalism and the Good for a Person". *Ethics*, 106, 297-326.
- Schroeder, T. (2004). *Three Faces of Desire*. Oxford: Oxford University Press.
- Sidgwick, H. (1981). *The Methods of Ethics*. Indianapolis: Hackett.

- Silverstein, M. (2000). "In Defense of Happiness: A Response to the Experience Machine". *Social Theory and Practice*, 26, 279–300.
- Skow, B. (2009). "Preferentism and the Paradox of Desire". *Journal of Ethics and Social Philosophy*, 3, 1-17.
- Sobel, D. (2002). "Varieties of Hedonism". *The Journal of Social Philosophy*, 33, 240-256.
- Sober, E. (2001.) "What is the Problem of Simplicity?" In Zellner et al. (eds.) *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*, pp. 13-31. Cambridge: Cambridge University Press.
- Sober, E. (2013). "Psychological Egoism". In H. LaFollette and I. Persson eds. *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell.
- Sumner, L.W. (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- Swinburne, R. (1997). *Simplicity as Evidence of Truth*. Milwaukee: Marquette University Press.
- Tännsjö, T. (1998). *Hedonistic Utilitarianism*. Edinburgh: Edinburgh University Press.
- Velleman, D. (1991). "Well-Being and Time". *Pacific Philosophical Quarterly*, 72, 48-77.
- Velleman, D. (1998). "Is Motivation Internal to Value?" In C. Fehige & U. Wessels (eds.), *Preferences*. Berlin: Walter de Gruyter.
- Weijers, D. (2014). "Nozick's Experience Machine is Dead, Long Live the Experience Machine!" *Philosophical Psychology*, 27, 513-535.
- Wittgenstein, L. (1973). *Philosophical Investigations*. Ed. and trans. G.E.M. Anscombe. Upper Saddle River, New Jersey: Prentice Hall.