University of Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Arts and Humanities

Department of Philosophy

A Defence of the Desire Theory of Well-being

by

Atus Mariqueo-Russell

ORCID ID 0000-0002-6640-322X

Thesis for the degree of Doctor of Philosophy

August 2023

University of Southampton

Abstract

Faculty of Arts and Humanities

Department of Philosophy

Thesis for the degree of Doctor of Philosophy

A Defence of the Desire Theory of Well-being

by

Atus Mariqueo-Russell

Desire theories of well-being claim that how well someone's life goes for them is entirely determined by the fulfilment and frustration of their desires. This thesis considers the viability of theories of this sort. It examines a series of objections that threaten to undermine these views. These objections claim that desire theories of well-being are incorrect because they have implausible implications. I consider four main objections over the course of this thesis. The first claims that these theories are incorrect because they implausibly entail that self-sacrifice does not exist. The second claims that these theories are incorrect because they implausibly entail that severe depression does not diminish the well-being of those afflicted by this condition. The third claims that these theories are incorrect because they have implausible implications about the relative importance of fleeting desires, long-standing desires, and fluctuations in desire strength to well-being. The fourth claims that these theories are incorrect because they fail to capture the intuition that desire fulfilments which leave us disappointed and bereft of feelings of satisfaction do not improve well-being. In each of these cases, I find that desire theories of well-being have sufficient resources to refute these objections. The primary finding of this thesis is that many of the arguments against desire theories of well-being are unsuccessful. A secondary set of findings concern observations about the structure of human psychology.

Keywords: Dead Sea apples, Depression, Desire Satisfactionism, Motivation, Preferentism, Prudential reasons, Self-sacrifice, Unstable desires, Welfare.

Table of Contents

Table of Figures

Research Thesis: Declaration of Authorship

Print name: Atus Mariqueo-Russell

Title of thesis: A Defence of the Desire Theory of Well-being

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Mariqueo-Russell, A. (2023). Desire and motivation in desire theories of well-being. *Philosophical Studies*, 180(7), 1975–1994. https://doi.org/10.1007/s11098-023-01966-y

and

Mariqueo-Russell, A. (2023). Well-being and the problem of unstable desires. *Utilitas*, 35(4), 260–276. https://doi.org/10.1017/S0953820823000171

Signature: ..... Date: 29/08/2023

Acknowledgements

Thanks are due first and foremost to my supervisors Alex Gregory and Brian McElwee who have painstakingly commented on multiple drafts of this thesis.

I would also like to thank the following people for comments on drafts of material that found its way into individual chapters: Giulia Felappi, Charlotte Franziska Unruh, Francesco Gandellini, Mariana Guerra, St.John Lambert, Maria Mourtou Paradeisopoulou, Harish Narayanan, Elliot Porter, Andrew Spaid, Perrine Sriwannawit, Kurt Sylvan, and Jonathan Way. Thanks also to three anonymous referees from the journal *Philosophical Studies*, two from *Utilitas*, and to the editor of *Utilitas*, Ben Eggleston.

This thesis has been immeasurably improved by the philosophy community at the University of Southampton, and particularly by audiences at four PGR seminars where drafts of this material were presented. Chapter Three has also benefitted from audience feedback from the 2022 PhiGS Colloquium at the University of Kent.

Finally, this research was made possible by generous funding provided by the University of Southampton's Presidential Scholarship, and by the Royal Institute of Philosophy's Jacobsen Scholarship.

Introduction

The term ‘well-being’ describes the value that tracks how well a life goes for the subject living it. Desire theories of well-being are a family of views about what determines this value. While there are many different versions of this theory, they coalesce around the following two claims:

1. Only the fulfilment of a subject’s (well-being relevant) desires non-instrumentally increases their well-being, while only the frustration of their (well-being relevant) desires non-instrumentally decreases their well-being.

2. The extent to which a subject is made better or worse off by a desire fulfilment or frustration is proportional to the strength of their fulfilled or frustrated desire.

This thesis considers the viability of theories of this sort. It examines a series of objections that threaten to undermine these views. These claim that desire theories of well-being are incorrect because they have implausible implications. I consider four main objections over the course of this thesis. The thesis is structured in the following way:

- Chapter One gives an overview of the state of the literature on well-being, the methodological approach used, and the background assumptions of this discussion.
- Chapter Two considers the objection that desire theories of well-being are incorrect because they implausibly entail that self-sacrifice does not exist.
- Chapter Three considers the objection that desire theories of well-being are incorrect because they implausibly entail that severe depression does not diminish the well-being of those afflicted by this condition.
- Chapter Four considers the objection that desire theories of well-being are incorrect because they have implausible implications about the relative importance of fleeting desires, long-standing desires, and fluctuations in desire strength to well-being.
- Chapter Five considers the objection that desire theories of well-being are incorrect because they fail to capture the intuition that desire fulfilments which leave us disappointed and bereft of feelings of satisfaction do not improve well-being.

In each of these cases, I find that desire theories of well-being have sufficient resources to refute these objections. Consequently, I argue that these theories ought not to be rejected on the basis of these problems. The primary finding of this thesis is that many of the arguments against desire theories of well-being are unsuccessful. A secondary set of findings concern observations about the structure of human psychology.

There are broadly two types of responses that proponents of desire theories of well-being typically give when confronted with objections of this sort. The first involves modifying the theory so that it no longer has the counterintuitive implications that are said to undermine it. The second involves denying the force of objections that are said to undermine simpler versions of the theory. Often proponents of this second approach point to other plausible philosophical commitments that explain why these theories do not lead to the counterintuitive implications that the objector claims, or at least why these implications are not so counterintuitive as to warrant rejecting the theory. I principally make arguments of this second type. I argue that the problems considered in Chapters Two, Three, and Five are all navigable by making plausible and independently motivated claims about the structure of human psychology. Conversely, Chapter Four relies upon an argument of the first type. In that chapter, I argue that desire theories of well-being must welcome some additional complexity if they are to remain viable in the face of the problem of unstable desires.

Over the course of this thesis, I commit to several claims about the structure of human psychology. There are three in particular that contribute to solutions across several of the problems considered. These are: that we have some basic standing desires and aversions for certain affective states (§1.3; §3.4; §4.3; §5.4), that desires do not always motivate proportionally to their strength (§2.3; §3.2; §3.4), and that some desires are conjunctive (§3.3; §5.3; §5.4). I highlight these claims here because, while each is independently plausible, they are not widely appealed to in response to objections to desire theories of well-being. If any of these claims are incorrect, then several of my arguments will need to be reworked or rejected. Nevertheless, I am confident that these claims are robust.

Some of the findings of this research have been published elsewhere during the course of writing this thesis. Chapters Two and Three both develop material present in my paper ‘Desire and motivation in desire theories of well-being’ published in *Philosophical Studies* (Mariqueo-Russell 2023a). These chapters expand upon the ideas put forward in this article in the following ways. Chapter Two considers and rejects an additional response to the problem of self-sacrifice (§2.2.2). This chapter also contains a more in-depth discussion of the claims that I make about moral psychology (§2.4.1–§2.4.3). Chapter Three provides supplementary arguments for how desire theories of well-being can account for the harm of residual cases of severe depression (§3.4). This chapter also contains a new discussion about the relationship between severe depression and prudential reasons (§3.5). None of the material in these chapters contradicts the arguments found in this paper.

The argument put forward in Chapter Four builds upon material present in my article 'Well-being and the problem of unstable desires' published in *Utilitas* (Mariqueo-Russell 2023b). This chapter contains an additional section which considers how these findings affect our wider understanding of prudential reasons and certain mental disorders (§4.5). Chapter Five expands upon this argument by considering how the view defended in Chapter Four ought to account for minor changes in the content of desires (§5.4). Aside from this, the substance of Chapter Four remains similar to that found in the paper. Again, none of the material in this thesis contradicts the arguments made in this paper.

This thesis contributes to the literature on well-being by rebutting several important objections to desire theories of well-being. My arguments do not establish that desire theories of well-being are the correct, or even most plausible, family of views about well-being. Indeed, there remain important objections that proponents of these theories need to solve. Nevertheless, this thesis shows that several influential objections to these theories are navigable. Therefore, those who want to reject these theories are better off appealing to other objections or making the positive case for their preferred alternative.

Chapter 1 Desire Theories of Well-being

1.0 Introduction

This thesis argues that certain versions of the desire theory of well-being can overcome a series of objections that threaten to undermine this family of views. Over the course of the chapters that follow, I outline the philosophical commitments that proponents of these theories ought to make in order to solve these problems. If this project is successful, then opponents of desire theories of well-being will be deprived of several influential arguments against these views. I do not claim that a particular version of the desire theory of well-being is the correct theory of well-being. Nor do I make metaethical commitments about what it means to call a theory of well-being correct. Nevertheless, my aspiration is for readers previously unconvinced by these theories to reconsider their viability in light of these findings. This chapter provides an overview of desire theories of well-being for readers unfamiliar with the contours of debates about well-being within analytic philosophy. This sets the scene for my later discussion of objections that threaten to undermine these theories.

This chapter has the following structure: §1.1 begins by outlining the concept of well-being and explains my approach to this topic. §1.2 outlines a minimally plausible version of the desire theory of well-being and makes the case for finding this position initially attractive. §1.3 examines mental state theories of well-being and gives reasons for rejecting these views in favour of desire theories of well-being. §1.4 repeats this approach with objective list theories. §1.5 concludes with a summary.

1.1 The concept of well-being

The term ‘well-being’ describes the value that determines how well a life goes for the subject living it. Well-being plays a key role in several important normative frameworks. For instance, it is the concept that we employ whenever we consider what is in someone’s best interests. It is the value at the centre of theories of prudence, the virtue of beneficence, and utilitarian theories of ethics. It is the central value expressed in what we want for others insofar as we care for them (Darwall 2002). It is the value that we are concerned with when we consider what someone deserves in terms of reward or punishment, and when we feel sorry or glad for someone (Pallies 2022, 600). Care for a loved one’s well-being also seems

to be a central part of plausible theories of love (Frankfurt 2004). Consequently, our theory of well-being has ramifications for a whole host of our other normative perspectives.

It may turn out that the content of well-being is entirely determined by the presence, absence, and inverse of pleasure, happiness, desire fulfilment, goal achievement, virtuous activity, life satisfaction, certain emotional states, some combination of these, or something else entirely. However, the concept of well-being is distinct from candidate theories about its content. This distinction is what allows debates about the constituent parts of well-being to be meaningful. Some writers prefer to use the terms 'welfare', 'flourishing' or 'prudential value' instead of 'well-being'. One advantage of these terms is that they have fewer connotations of serene or gratifying mental states. In everyday language, and academic fields outside of philosophy, the term 'well-being' is frequently used to refer to a cluster of phenomenologically pleasant mental states. Ideally our terminology for the value under consideration should remain neutral about the content of that value. Using the term 'well-being' risks prejudicing readers in favour of mental state views. Consequently, there is something to be said for alternative terminologies.

Nevertheless, rival terms also have significant downsides. The term 'welfare' is relatively neutral as to the content of the value under consideration. However, in everyday language it is often used to refer to social security payments. This makes it potentially misleading to prospective readers. The term 'flourishing' does not have strong connotations of mental states. However, it is an often-used translation of the Ancient Greek word 'eudaimonia'. This term involves specific theoretical commitments as to the content of this value (§1.4). 'Prudential value' is probably best at remaining neutral between different theories and precise about what is under discussion. However, this term has the unfortunate downside of being unrecognisable to anyone outside of analytic philosophy. Contributing to the incomprehensibility of analytic philosophy is too great a burden for my conscience to bear. Consequently, I have opted to use the term 'well-being' instead. Nevertheless, if readers prefer an alternative, then they are free to substitute 'well-being' with 'welfare', 'flourishing' or 'prudential value' as they read this thesis.

Most discussions about well-being within analytic philosophy focus on its normative dimension. Normative theories aspire to provide an account of the principles that determine well-being. These principles allow us to evaluate whether particular states of affairs are non-instrumentally good or bad for someone. Analytic philosophy has tended to pay less attention to applied questions about which particular states of affairs generally improve or decrease

well-being. It has been even more quiet on metaethical questions about whether theories of well-being are the sort of thing that can be objectively correct.

This thesis considers a normative theory of well-being. It avoids giving practical life advice or becoming mired in metaethical debates. However, it cannot remain completely neutral on the metaethics of well-being. This is because some metaethical views are incompatible with fruitful normative investigation. For instance, error theory claims that all normative statements that postulate the existence of well-being are simply incorrect (Mackie 1977). If this type of view is right, then investigations into normative theories of well-being are made obsolete because all theories that postulate the existence of well-being are equally incorrect. Consequently, while this thesis can remain neutral on a wide range of metaethical positions, it must reject the error theory if it is to avoid redundancy. Aside from this caveat, I do not assume a particular position on the metaethics of well-being in this thesis.

Normative theories of well-being seek to identify the principles that determine how well a life goes for the subject living it. Most such theories also provide an account of ill-being. This term refers to the value that determines what makes a life go badly for the subject living it (Kagan 2014). Ill-being is often said to be constituted by either the absence or inverse of the properties that improve well-being. For instance, hedonism claims that pleasure improves well-being, while pain increases ill-being (§1.3). Well-being and ill-being are commonly conceptualised as tracking the same value. On this view, 'ill-being' is simply a descriptor for negative well-being, and 'well-being' is a descriptor for negative ill-being.

Some writers think that ill-being is a distinct value to well-being. If they are right, then it may be that ill-being is incommensurable or incomparable with well-being. Ruth Chang defines two things as incommensurable when they do not share properties that allow them to be precisely compared to a single scale of measurement (1997, 2). For instance, the aesthetic value of preserving the works of Shakespeare is incommensurable with the ethical value of saving ten thousand lives if we cannot make precise cardinal judgements about their relative value. She contrasts the notion of incommensurability with the stronger claim of incomparability. Two things are incomparable when they cannot be non-arbitrarily compared to a single scale of measurement because no positive value relation holds between them (Chang 1997, 4). Whereas incommensurability allows for non-precise comparisons, incomparability entails that no such comparisons can be non-arbitrarily made. The incommensurability or incomparability of values may go some way towards explaining how moral and tragic dilemmas arise (Richardson 1994, 115–117; Nussbaum 1986, 47). In such

cases, choices are dilemmas because they stand to actualise different values in such a way that some outcomes are neither better than, worse than, nor equal to each other.

Some people think that values are not the sort of thing that can be incomparable with each other. This view is supported by the role that deliberation plays when values conflict. When prospective mutually exclusive actions result in the realisation of different values, we can deliberate to try to ascertain which outcome is better. Conversely, deliberation seems impossible in non-normative cases of incomparability (Kelly 2008, 372). The question of whether a room is more or less hot than it is humid typically invites exasperation. Conversely, the question of whether one ought to donate £100 to charity or spend it on oneself provokes deliberation about the potential trade-offs between the ethical and prudential values at stake. While it is not always clear what the correct decision is, there nevertheless seems to be a process of arriving at decisions when values are concerned. This suggests that, while it can sometimes be difficult to identify the relative weight of different values, they are nevertheless comparable.

Regardless of our position on these issues in general, there are independent reasons to think that well-being and ill-being track the same specific underlying value. The trade-offs that we make between them are relatively straightforward when compared to those made between other values. Moreover, it is difficult to postulate compelling moral or tragic dilemmas to illustrate incomparability or incommensurability when these are the only two values at stake. This suggests that they are not distinct values. Furthermore, if they were distinct values, then an individual could have high well-being *and* high ill-being (Tully 2017, 3–4). This is a conclusion that we should be reluctant to accept. It would mean that the all-things-considered judgements that we make about how well a life has gone for someone are neglectful of the fact that such judgements are comprised of two separate values. Consequently, we should assume that the terms ‘well-being’ and ‘ill-being’ track the same underlying value unless we encounter something in our investigation that suggests otherwise. If this is right, then we can set aside the threats that incommensurability and incomparability pose to this analysis.

Desire theories of well-being are also theories of ill-being.[1] They purport to tell us what makes a life go better, what makes it go worse, and by how much. These views claim that the terms ‘well-being’ and ‘ill-being’ refer to the same value. If desire theories of well-being

[1] Shelly Kagan argues that desire theories of well-being are made more plausible when applied solely to well-being’s positive dimension (2014). Nevertheless, most proponents of these theories consider them to also account for ill-being. I assume the standard view in this thesis.

are unable to capture intuitions about ill-being, then there is cause to reject these theories, or to restrict them to only accounting for well-being's positive dimension. However, in the absence of successful counterarguments, I shall proceed by taking these theories to account for both well-being and ill-being. For ease of reading, I will generally refer to increases and decreases in well-being and largely omit further mention of ill-being.

There remains a methodological question about how normative theories ought to go about identifying the principles that determine well-being. Often discussions about well-being do not specify a particular methodological approach. Nevertheless, the vast bulk of these discussions implicitly, and sometimes explicitly, appeal to reflective equilibrium. This approach involves adjusting our pre-theoretical intuitions about well-being until we reach a point when our normative theory entails a coherent set of valuations. It would be too high a burden to place on any theory to expect it to perfectly capture all of our pre-theoretical intuitions about well-being. Indeed, these intuitions are often inconsistent and contradictory. Reflective equilibrium requires a willingness to revise intuitions that do not align well with our wider system if that system can capture our intuitions better than its competitors. This is probably the dominant approach within the literature on well-being, and the one applied in this thesis. Consequently, this thesis seeks to establish that a version of the desire theory of well-being entails a coherent set of valuations that captures the bulk of our intuitions about well-being – at least in relation to the problems considered.

There seem to be at least three types of intuitions that are relevant to the evaluation of theories of well-being. The first is extensional adequacy. This concerns the extent to which a theory has intuitive implications about which specific things improve or reduce well-being. For instance, if a theory entails that severe depression does not reduce well-being (Chapter Three), then this theory ought to be rejected for being extensionally inadequate. The second set of intuitions that is relevant to the evaluation of theories of well-being are those about descriptive adequacy (Sumner 1996, 10). This concerns the extent to which a theory is able to capture intuitions about the principles that determine well-being. For instance, if a theory requires a succession of arbitrary conjuncts and disjuncts in order to capture wider intuitions about well-being, then this theory ought to be rejected for being descriptively inadequate (Enoch 2005, 766–769). The final set of intuitions relevant to this investigation are those about normative adequacy (Sumner 1996, 8). This concerns the extent to which a theory coheres with our wider set of normative commitments. All three types of intuition should be factored into the reflective equilibrium process.

1.2 Desire theories of well-being

Desire theories of well-being are a family of views.[2] The most plausible versions of these theories contain variations of the following two features:

1. Only the fulfilment of a subject's (well-being relevant) desires non-instrumentally increases their well-being, while only the frustration of their (well-being relevant) desires non-instrumentally decreases their well-being.

2. The extent to which a subject is made better or worse off by a desire fulfilment or frustration is proportional to the strength of their fulfilled or frustrated desire.

This formulation constitutes a minimally plausible version of the desire theory of well-being. I take these features to be relatively uncontroversial among contemporary proponents of these theories. Nevertheless, there are dissenting views that may also lay claim to this nomenclature. It is possible to broaden this formulation to include these views. However, doing so risks making the features that the most plausible versions of these theories share less salient. If desire theories of well-being are a family of views, then we can consider theories that bear resemblance to, but deviate from, these features to be estranged cousins.[3]

The minimally plausible desire theory of well-being is imprecise. The first feature is strategically ambiguous about which desires are relevant to well-being. This allows it to accommodate differing opinions about whether we ought to restrict the subset of desires that determine well-being and, if so, how we ought to restrict them. The problems considered in this thesis can be navigated without restrictions of this sort. Therefore, these arguments are compatible with the view that all desires are relevant to well-being (Lukas 2010). If this view is correct, then our formulation can do without the bracketed clauses. Nevertheless, it may be that some restrictions are warranted in response to alternative problems or are attractive in and of themselves.[4] Consequently, this thesis remains neutral about which desires are relevant to well-being.

---

[2] These theories are sometimes referred to by other names, such as: Desire satisfactionism, desire-satisfaction theories, desire-fulfilment theories, desire-based theories, desire accounts, and preferentism.

[3] For instance, antifrustrationism resembles desire theories of well-being despite rejecting the first feature. This view accepts that only desire frustrations decrease well-being but claims that nothing increases well-being (Fehige 1998).

[4] For example, some writers argue that desire theories of well-being should exclude instrumental desires from non-instrumentally affecting well-being (Sarch 2013, 223; Heathwood 2005, 489; Brandt 1979, 111). I discuss this view in §5.2.3. Other proposed restrictions are considered in §2.2.1 and §5.2.2.

The second feature of the minimally plausible desire theory of well-being does not specify at which time the strength of desire modifies the extent of its effects on well-being. Consequently, this theory cannot be operationalised until the second feature is altered to address this deficit. It is possible to quibble that calling this view 'minimally plausible' is a misnomer given that it is unusable without further elaboration. Nevertheless, I favour this terminology because modified versions of the second feature are likely to be compatible with the second feature as it stands. This is because they are likely to simply further specify how desire strength should be integrated into the framework. Chapter Four argues for a specification of this sort. Nevertheless, for now I will assume that features 1 and 2 constitute a good starting point for a minimally plausible version of the desire theory of well-being.[5]

Sometimes desires can be fulfilled or frustrated without producing feelings of satisfaction or frustration. Consider the case of an exile who is cut off from information about the lives of their children (Parfit 1984, 495). Assume that the exile retains a strong desire for their children to live good lives, while lacking any information about them. Whether their children do in fact live good lives is what determines whether their desire is fulfilled or frustrated irrespective of whether they learn about their children's lives. Desire theories of well-being entail that whether the exile's desire is fulfilled or frustrated affects their well-being despite its failure to produce any beliefs or feelings within them.

These theories also capture the intuition that misperceptions and deceptions are less conducive to well-being than phenomenologically identical veridical perceptions. Robert Nozick invites us to imagine a machine that simulates experiences in a way that is perceptually indistinguishable from reality. The machine is programmed to be optimal at generating positive affective states through its simulations (Nozick 1974, 42–43). Proponents of desire theories of well-being can point out that, despite appearances to the contrary, many of our desires are frustrated in the machine. This is because we are often not actually getting what we want when we mistake its simulations for reality.[6] For instance,

[5] This formulation remains neutral between the 'combo' and 'object' versions of the desire theory of well-being. The 'combo' view claims that well-being is improved when a desire corresponds to a state of the world in which it is fulfilled, and well-being is decreased when a desire corresponds to a state of the world in which it is frustrated (Bradley 2007, 47; Kagan 1994, 312). Conversely, the 'object' view claims that it is the states of affairs that we desire that are non-instrumentally valuable, rather than relational states of desire and world (van Weelden 2019; Dorsey 2012a, 272–274; Dorsey 2013, 152–153).

[6] While this is frequently cited as a virtue of desire theories of well-being, it may be that these views do not fare much better than hedonism when it comes to the experience machine (Lowe & Stenberg 2017). This is because, while some desires are frustrated in the machine, it is possible that more

desires for things such as true beliefs are necessarily frustrated by the illusions that the machine produces. Moreover, relational desires for things such as friendship cannot be fulfilled by simulations. This is because these desires specify the existence of other persons in their content. According to desire theories of well-being, our well-being is not solely determined by our mental states. This makes these theories substantially different from the mental state theories of well-being considered in the next section.[7]

I have referred to desire theories of well-being as a family of views. Like most families, there is a good deal of resemblance between its members. Nevertheless, there are also relatives with significantly different features. This diversity makes it difficult to make a general case in favour of these views. Some attractive features of the majority of these theories do not apply to a smaller subset, and some virtues of individual views are not shared by the bulk. Consequently, I proceed by outlining the merits of what I take to be the most compelling formulations of these theories. Where helpful I caveat my exposition by noting how some points do not apply to certain versions of the theory. Nevertheless, this process of caveating does not exhaust the conceptual space of what it is possible for a desire theory of well-being to claim. Such an endeavour would reduce much of this discussion to a litany of overburdened footnotes. Even then it would fail to account for all possible variations.

There are several reasons why desire theories of well-being are attractive. Probably most prominent among them is the ability of these theories to capture the resonance requirement (Dorsey 2013, 157).[8] This is the intuition that our well-being must be something that we ordinarily find compelling and attractive, or at least would do if we were more rational and well-informed (Railton 1986, 9). The idea that what makes your life go better for you must

---

desires are fulfilled than they would be outside of it. For instance, this may be true of our sensorial desires. Perhaps the fulfilment of such desires outweighs the frustrations incurred in the machine.

[7] This analysis excludes Chris Heathwood's Subjective Desire Satisfactionism from the family of desire theories of well-being. Heathwood has argued that well-being is improved by the subjective perception that a desire has been fulfilled and decreased by the subjective perception that a desire has been frustrated (2006, 559). L. W. Sumner points out that desire theories of well-being that integrate an experiential component inevitably end up 'mutating into an experiential theory with no desire component' (1996, 128). This seems true of Subjective Desire Satisfactionism, which Heathwood claims is extensionally equivalent to attitudinal hedonism (2006, 539). This is one of the mental state theories of well-being considered in §1.3.

[8] A related desideratum is existence internalism (Dorsey 2012a; Noggle 1999, 303; Rosati 1996; Rosati 1995, 300; Loeb 1995). This is the view that our own well-being must be something that we are ordinarily motivated to pursue, or at least would be if we were more rational or informed. Existence internalism is a narrower version of the resonance requirement (Dorsey 2012a, 275). It requires that the pursuit of our own well-being motivates, rather than simply resonates. I focus on the resonance requirement in order to spare readers from arduous repetition.

be something that resonates with some of your own pro-attitudes is compelling. However, despite its plausibility, many theories of well-being fail to capture this requirement. There are a limited number of candidates for well-being goods that resonate with all well-being subjects. The fulfilment of our desires is one such candidate. There are mental states, such as happiness or pleasure, that are also candidates.

Some versions of the desire theory of well-being fail to capture the resonance requirement. This is true of some idealisation theories. These theories claim that it is not the fulfilment and frustration of our actual desires that determines our well-being, but rather that of the counterfactual desires that we would have in an idealised set of circumstances. For instance, Henry Sidgwick considers the view that a person's good is 'what would be desired, with strength proportioned to the degree of desirability, if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition' (1907, 111). This type of theory runs the risk of postulating an alienated standard of well-being. After all, the desires that we would have if we were to undergo a radical transformation may fail to resonate with us as we actually are (Rosati 1995, 309–311).

Whether these views fail to capture the resonance requirement depends upon how broadly we interpret its proviso. It is possible to claim that a perfectly rational and well-informed person would have a completely different set of subjective attitudes than they actually have. If this is right, then it seems that most theories of well-being can capture the resonance requirement. These theories simply need to claim that their standard of well-being would resonate with perfectly rational and well-informed beings. However, this type of speculative manoeuvre undermines the impetus behind the resonance requirement. If every theory of well-being can capture it, then it offers no way to discern convincing from unconvincing theories. Consequently, the resonance requirement needs to be refined so that counterfactual desires of this sort do not satisfy it. Connie Rosati claims that we should require that any idealisation incorporates only those conditions that can be said to bear a reasonable relationship to the concerns that animate inquiry into well-being (1995, 301). This rules out radical forms of idealisation (§5.2.2).

Some types of idealisation theory can make a plausible claim to capture the resonance requirement. Dale Dorsey argues for a theory of this sort (2021, ch.6). On his view, we ought to count the things that we actually value as determining our well-being but idealise our other pro-attitudes to make them align with these values (Dorsey 2021, 143). We are not alienated from these counterfactual pro-attitudes because they still resonate with our values. Dorsey

advances a project-based subjectivist view rather than a desire theory of well-being. Nevertheless, subjecting desires to a process of idealisation grounded in other subjective attitudes is something that proponents of desire theories of well-being could emulate. For instance, we could idealise our instrumental desires in light of our intrinsic desires. Alternatively, we could idealise our first-order desires in light of our second-order desires; or a subset of our intrinsic desires in light of another subset of our intrinsic desires. Idealisation theories of this type are compatible with the resonance requirement because the things that they claim improve our well-being are things that resonate with our existent pro-attitudes.

Idealisation aside, there are other types of desire theory of well-being that also fail to capture the resonance requirement. This is true of some versions of the theory that reject concurrentism (§5.2.1). This is the view that desires must exist at the same time as the states of affairs specified in their content in order for their fulfilment or frustration to affect well-being (Heathwood 2005, 490).[9] Concurrentism entails that if a state of affairs comes about after we have ceased to desire it, then it does not improve well-being. Views that reject concurrentism allow lost desires to sometimes improve well-being if the content of those desires occurs after the desire ceases (Bruckner 2013; Dorsey 2013; Vorobej 1998). This approach is often appealed to in order to account for posthumous benefits and harms (Pitcher 1984). Some views that reject concurrentism are committed to an alienated standard of well-being. This is because lost desires often fail to resonate with us.[10]

Whether a particular version of the desire theory of well-being captures the resonance requirement depends upon the point in time that it claims that desire fulfilments and frustrations affect well-being. The time-of-object view claims that well-being is affected at the time that a desired state of affairs occurs or fails to occur (Dorsey 2013, 156). This can be after we have ceased to have the desire. The time-of-object view is committed to an alienated standard of well-being (Dorsey 2013, 156–157). This is because it claims that the

[9] There are different views that go by the name concurrentism. Another such view claims that the extent of the effects that a desire fulfilment or frustration has on well-being is determined by the strength of the desire at the time of its fulfilment or frustration. This position is considered in Chapter Four.

[10] This conclusion depends on how exactly we formulate the resonance requirement. In this discussion I am assuming Dorsey's view that well-being goods must resonate with us at the time of their acquisition in order for our theory to capture the resonance requirement (2013, 156–157). However, it is possible to reject this idea and instead claim that providing a well-being good did at some point resonate with us, then its acquisition captures the resonance requirement (Lin 2017b, 179–180).

acquisition of things that fail to resonate with us can improve our well-being. However, not all views that reject concurrentism are committed to the time-of-object view. For instance, the time-of-desire view claims that well-being is affected at the time that we have the desire (Dorsey 2013, 158–159). This can be prior to the desired state of affairs occurring or failing to occur. On this view, fulfilling a lost desire improves well-being at the time that that desire was held. Consequently, this view captures the resonance requirement because the things that improve well-being always resonate with us at the time that they improve well-being.[11]

A second reason to find desire theories of well-being attractive is for their conceptual parsimoniousness. All things being equal, simpler theories of well-being are better placed to placate accusations of arbitrariness than their more complex competitors. Such theories typically have a higher degree of descriptive adequacy. The conceptual parsimoniousness of desire theories of well-being comes in different degrees depending upon the version of the theory under consideration. For instance, idealisation theories are significantly less conceptually parsimonious than views that take our actual desires to determine our well-being. In the chapters that follow I put forward a version of the desire theory of well-being that seeks to preserve as much conceptual parsimoniousness as possible while successfully navigating the problems considered. Of course, there are other theories of well-being that can also lay claim to this virtue. Mental state theories of well-being of the sort considered in the next section are often equally conceptually parsimonious.

A third promising feature of desire theories of well-being is their ability to postulate a universal standard of well-being. These theories can claim that any being that is capable of having desires is a subject capable of well-being. This allows us to identify who qualifies as a well-being subject as well as how to work out their level of well-being (Sumner 1996, 123). Consequently, the explanatory value of these theories surpasses that of rival views that purport to explain only human well-being. Rival theories that require species-specific standards of well-being may struggle to explain how to non-arbitrarily identify which standards apply to which species (Lin 2018, 324). Not only can the claims that these views make about human well-being be contested, but the criteria that they use to work out which standards apply to which species can also be challenged. This heightens their vulnerability to criticism. Desire theories of well-being are not alone in their ability to postulate a universal standard of well-being. Simple mental state theories of well-being, such as hedonism, are

[11] There are alternative views about the time at which desires improve well-being. For instance, Eden Lin argues that sometimes well-being is affected at the time of object, and sometimes at the time of desire (2017b). I set aside such views here.

also well-placed to make this claim. Moreover, it is possible to restrict which species desire theories of well-being apply to if there are independent reasons for doing so. Nevertheless, the fact that these theories can be universally applied makes them more compelling.

A fourth attraction of desire theories of well-being is that these views are operationalisable in policymaking. For instance, economists often tacitly, and sometimes explicitly, appeal to desire theories of well-being when articulating the importance of factoring preferences into their evaluations of policy proposals. The measurement and valuation of preferences has become an increasingly central part of the process of cost-benefit analysis (Sunstein 2018). Desire theories of well-being seem to fit relatively well within this evaluative approach. Perhaps practical applicability is not a good reason to think that these theories are any more likely to be objectively correct than their rivals. Nevertheless, the fact that this view can be operationalised within policymaking may be particularly relevant if it turns out that no theory of well-being is objectively correct, or if multiple theories of well-being turn out to be correct. Were that to be the case, then the practical applicability of theories of well-being is likely to be an important consideration when comparing rival views. It is also worth noting that other theories of well-being can also make viable claims to being operationalisable in policymaking. Moreover, the number of theories able to do so is likely to proliferate as the science of well-being measurement advances (Alexandrova 2017).

A fifth reason for finding desire theories of well-being attractive is because these views fit well with certain strains of liberal political philosophy. Both political liberalism and desire theories of well-being place an emphasis on the sovereignty of the individual in determining their own conditions for living well. A normative theory of well-being may be more attractive if it does not require us to revise our intuitions in domains such as political theory. Given that liberalism is still—just about—the dominant political philosophy of our age, the fact that desire theories of well-being coexist with it in a relatively straightforward and frictionless manner is sometimes taken to be a point in their favour (Sumner 1996, 123). Of course, not everyone is a liberal. If we are looking for a theory of well-being to cohere with non-liberal politics, then the perceived proximity of desire theories of well-being to liberalism may count against these theories.

I began this section by outlining desire theories of well-being. I then put forward five reasons to find these views initially compelling. These are: their ability to capture the resonance requirement, their ability to be conceptually parsimonious, their ability to postulate a universal standard of well-being, their operationalisability in policymaking, and their compatibility with liberal political philosophy. I have outlined these reasons in what I

consider to be their descending order of importance. This is also broadly reflective of the amount of attention that each has received in the literature on well-being. In the next section I consider some mental state theories of well-being that are equally well-placed to capture the first three attractions outlined here. I argue that, despite sharing these attractions with desire theories of well-being, these theories incur additional costs that make them worse starting points for understanding well-being.

1.3 Against mental state theories of well-being

A common family of views claims that well-being is solely determined by the possession and absence of certain mental states. These are mental state theories of well-being, or simply 'mental state theories' for short. One such mental state theory is hedonism. This is the view that well-being is solely determined by pleasure and pain. According to hedonism, experiences of pleasure increase well-being, while experiences of pain decrease well-being. On this view, gains and losses of well-being depend upon the duration and intensity of pleasures and pains, with longer and more intense feelings having larger effects on well-being (Feldman 2004, 25). Hedonist views typically also claim that a subject's lifetime well-being is solely determined by aggregating every instance of pleasure and pain, when adjusted for intensity, throughout the course of that life (Feldman 2004, ch.6).[12]

Hedonism is the most well-known mental state theory. However, among hedonists there is considerable disagreement about what constitutes pleasure and pain. Epicurus considered the absence of pain to be the definition of pleasure. Jeremy Bentham emphasised the sensory aspects of pleasure (1789). John Stuart Mill arguably considered intellectual pleasures to have lexical priority over sensory pleasures (1861). Whereas Fred Feldman considers pleasure to be a propositional attitude that is always, although not exclusively, held towards sensory pleasures (2004). It is also worth considering the possibility that pleasure may be a heterogeneous concept that encompasses various mental states. If this is the case, then questions emerge about the well-being weighting of different pleasurable mental states and

[12] It is possible to reject this view and instead claim that well-being is non-additive (Velleman 1991). On such views, the distributional shape of well-being throughout the course of a life is non-instrumentally relevant to the total well-being value of that life. I set aside this issue here.

about what unifies them into a single kind (Griffin 1986, 8).[13] Consequently, the term hedonism encompasses a rich variety of theories (Feldman 2004, 30).

Hedonism faces competitor mental state theories. For example, Subjective Desire Satisfactionism claims that well-being is solely determined by the feelings that emerge from subjectively perceiving that our desires have been satisfied or frustrated (Heathwood 2006).[14] Others have argued that feelings of happiness and unhappiness determine well-being (Almeder 2000). Conversely, broader experientialism claims that a range of mental states affect well-being (van der Deijl 2019). Indeed, many possible variations exist within and beyond these alternative mental state theories. Regardless of the specifics of the theory under consideration, all mental state theories are unified by the claim that well-being is solely determined by the presence and absence of mental states of some sort. They simply disagree about which types of mental states matter to well-being.

It is worth noting that mental state theories typically refer to a specific type of mental state: feelings. Views that take beliefs or desires to determine well-being are not mental state theories if they are concerned with the objective truth of those beliefs or the actual fulfilment of those desires. Such views are interested in the relationship between certain mental states and the external world. Consequently, they do not take the presence or absence of mental states to be solely responsible for well-being. As far as I am aware, no one has defended the view that certain beliefs or desires determine well-being irrespective of the truth of those beliefs or the fulfilment of those desires. If one were to hold such a view, then this would qualify as a mental state theory.

To my mind, there are at least four reasons why mental state theories are initially attractive. These concern the extent to which mental state theories capture intuitions about well-being. Of course, they do not apply equally to all mental state theories. Nevertheless, they are significant enough to warrant elaboration. The first three of these are shared by desire theories of well-being. The last is not.

---

[13] Several philosophers have argued that what grounds pleasures as a unified kind is that they share the same phenomenological quality (Bramble 2013; Smuts 2011; Crisp 2006a). Timothy Schroeder argues that pleasure is a unified kind and draws upon a mix of philosophical arguments and neuroscientific findings to support this claim (2004, 103–105). I set aside these questions.

[14] An alternative reading claims that it is the act of subjectively perceiving the satisfaction or frustration of our desires that determines well-being, rather than the feelings that emerge from this act. This alternative is better characterised as a variation of the desire theory of well-being, rather than a mental state theory.

Firstly, mental state theories can capture the resonance requirement. These theories claim that the things that improve well-being resonate with our own subjective pro-attitudes. Consequently, these theories do not postulate an alienated standard of well-being. Secondly, many mental state theories are remarkably conceptually parsimonious. This makes them less vulnerable to accusations of arbitrariness than their more complex competitors. Many of the most popular mental state theories claim that well-being is determined solely by two types of mental states, with one positively and one negatively affecting well-being. This may be true of hedonism, Subjective Desire Satisfactionism, and happiness theories. If a mental state theory identifies just two mental states as determining well-being, then its conceptual parsimoniousness may broaden its intuitive appeal. Of course, alternative mental state theories incorporate a plurality of mental states, such as those engendered by experiences of novelty, self-understanding, aesthetic values, and compassion (van der Deijl 2019, 1784). Broader experientialist accounts lack the same conceptual parsimoniousness. Thirdly, many mental state theories are well-placed to postulate a universal standard of well-being applicable to all subjects capable of well-being. This means that these theories are also able to identify who qualifies as a well-being subject. Mental state theories based upon complex feelings or compositions of feelings struggle to account for well-being in less complex creatures. Consequently, this attraction is forfeited by such theories. These three attractions of mental state theories are shared by popular versions of the desire theory of well-being.

The fourth attraction of mental state theories is that these theories conform to what James Griffin has labelled ‘the experience requirement’. This stipulates that the things that affect our well-being must enter into our experience (Griffin 1986, 16). While some people find this requirement independently intuitive, others adopt it as a response to the counterintuitive implications of non-mental state theories of well-being (Heathwood 2006; Griffin 1986). Desire theories of well-being do not capture the experience requirement. This is because these theories claim that desire fulfilments and frustrations affect our well-being even when we are unaware of whether our desire has been fulfilled or frustrated. Consequently, these states of affairs affect our well-being despite not entering into our experience.

These four features are not an exhaustive list of everything that we want from a viable theory of well-being. Moreover, mental state theories are not unique in their ability to capture these attractions individually. Rather, their strength is their capacity to meet them simultaneously. Consequently, these theories are popular partly because they fare well on these metrics.

There are many objections to mental state theories. Here I focus on a problem that often leads people to reject them in favour of a desire theory of well-being. This concerns the

intuition that the origin of mental states plays a role in determining their effects on well-being.

Many people have the intuition that mental states generated by misperception or deception do not contribute to well-being to the same extent as mental states with identical intrinsic properties that originate from true beliefs and perceptions. For the sake of linguistic brevity, I will call mental states generated by deception or misperception 'inauthentic' mental states, and those generated by true beliefs and perceptions 'authentic' mental states. Opinions diverge over how exactly inauthentic mental states should be treated. Some intuitions suggest that they contribute diminished amounts, nothing at all, or even reduce well-being. However, mental state theories typically treat them as having the same effects on well-being as authentic mental states with identical intrinsic properties. This is a problem for these theories.

Robert Nozick has a well-known thought experiment that highlights the intuitive force behind the view that inauthentic mental states affect well-being differently from their authentic counterparts (1974, 42–43). He points out that it is conceptually possible to imagine a machine that simulates experiences in a way that is perceptually indistinguishable from reality when we are plugged into it. The machine is programmed to be optimal at generating positive mental states through its simulations. It erases any memory we have of being plugged into it, and so prevents any distress that might arise from awareness of our situation. Consequently, it generates mental states with identical intrinsic properties to those that could be generated by a favourable reality. Mental state theories typically entail that a life plugged into the machine contains equal well-being to a life lived outside of it with authentic mental states that have identical intrinsic properties. This counterintuitive conclusion suggests that these theories are wrong.[15]

Some readers may be concerned that intuitions generated by such extreme hypothetical thought experiments are unreliable guides to axiology. However, Nozick's thought experiment can be easily reformulated in less extravagant terms. Shelly Kagan gives us one such reimagining. He asks us to imagine a deceived businessman. This businessman

[15] Some people do not share the intuition that a life lived within the experience machine contains less well-being than one lived outside of it with phenomenologically identical veridical experiences. Nevertheless, this does not necessarily militate in favour of hedonism. It may be that some people's desires are solely, or largely, for affective states (Baber 2010, 250). Consequently, desire theories of well-being entail that they are benefitted equally by illusions as they are by phenomenologically identical veridical experiences.

experiences many blissful mental states in response to beliefs that his life is going well. However, unbeknownst to him, he is deceived. His friends, colleagues, and family all hold him in deepest contempt and simply pretend to love, value, and respect him. The businessman's happiness is therefore completely contingent upon his false beliefs (Kagan 1994). We can compare the deceived businessman to his hypothetical twin, the undeceived businessman. The undeceived businessman experiences mental states with identical intrinsic properties to the deceived businessman. However, they originate from the true beliefs that his friends, colleagues, and family all love, value, and respect him. Intuitively, it seems that the former's well-being is at least marginally lower than that of the latter. I will refer to the inability of mental state theories to capture intuitions of this sort as 'the problem of inauthentic mental states'.

There are at least three possible arguments against the view that the inauthenticity of mental states non-instrumentally matters to well-being. Firstly, we could deny the intuitive force of the experience machine and deceived businessman examples. We could instead appeal to the widely used axiom that 'what you don't know can't hurt you'. Accepting this defuses the force of these examples. However, many people find such an approach to be counterintuitive enough to render mental state theories unviable.

Secondly, we could argue that our intuitions in these cases are mistaking lack of an alternative value for lack of well-being (Silverstein 2000). Fred Feldman proposes four values that are sometimes misidentified as well-being. These are: instrumental value, ethical value, aesthetic value, and exemplar value (Feldman 2004, 8–9). For example, it may be that a life lived in the experience machine lacks ethical value because it cannot display virtue, or it may lack aesthetic value if this cannot be optimally produced through simulation. However, I find this approach unconvincing. Many people have sufficiently fine-grained intuitions that radical deception is a paradigm case of harm. In my definition of well-being, I pointed out that well-being is at the heart of what we want for other people insofar as we care for them (§1.1). It is unusual for people to be indifferent about the radical deception or misperception of those that they care about. Consequently, it seems unlikely that we are mistaking lack of an alternative value for lack of well-being in these cases.

Thirdly, we could try and explain away intuitions about these cases by claiming that they are shaped by the fact that misperception and deception generally do instrumentally diminish well-being. Consequently, our intuitions may be ill-equipped to identify well-being in extreme cases where deception and misperception are guaranteed not to diminish well-being. According to this argument, an understanding of human psychology may help unfurl our

aversion to experience machine-type cases. However, this approach is unconvincing. It rests on the claim that we have misapprehended our own intuitions about well-being. If this were the case, then we should expect investigation into those intuitions to divest us of these misunderstandings. However, the intuitions that radical deception and misperception are non-instrumentally relevant to well-being are remarkably persistent in the face of such investigations. Consequently, we ought to take these intuitions seriously when considering the viability of theories of well-being.

The lives described in Nozick's experience machine and Kagan's deceived businessman examples are ones where their subjects are very rarely getting what they want. A solipsistic existence lived in the experience machine makes strong desires for things such as friendship impossible to fulfil. This is because these desires are premised upon the existence of other people. Consequently, mere simulacra will not fulfil them. It strikes me as implausible that a life where we consistently fail to get what we want is one that contains maximal well-being. While not everyone shares this intuition, it has proven forceful enough to be cited as a leading motivation for rejecting mental state theories (Baber 2010, 250; Lukas 2010, 6).

Some proponents of mental state theories take the problem of inauthentic mental states seriously. However, rather than abandon the mental state theory framework, they instead adopt a conditional mental state theory. Feldman articulates a position of this sort (2002). According to Feldman, while mental state theories require well-being to be dependent upon mental states, they can restrict which mental states improve well-being to only those that are based on correct beliefs and perceptions. On this view, certain mental states are necessary, but not sufficient, conditions for improving well-being. These theories make improvements to well-being conditional on things other than the intrinsic properties of mental states. Consequently, they allow facts about the origin of mental states to be non-instrumentally relevant to well-being. This allows these theories to claim that, under certain conditions, mental states that normally improve well-being fail to do so, or do so at a diminished rate, or even reduce well-being.[16]

[16] Roger Crisp makes a similar argument. He claims that the extent to which an episode of enjoyment benefits a subject is dependent upon how enjoyable that experience is, rather than simply how much it is enjoyed (Crisp 2006b, 110). Factors aside from mental states determine how enjoyable experiences are. This formulation allows us to claim that episodes of enjoyment that emerge from deception and misperception are, all things being equal, less enjoyable. They therefore benefit us less than veridical episodes of enjoyment. Wendy Donner has advanced a relevantly similar argument. She argues that qualitative hedonists can claim that the causal origins of pleasures influence the extent to which they affect well-being (Donner 1991, 72–75).

Feldman outlines a version of what he terms 'veridical intrinsic attitudinal hedonism' in response to the problem of inauthentic mental states (2002, 616). This theory claims that when pleasure is experienced in response to misperception or deception it affects well-being differently. This view can explain why someone envatted within the experience machine has less well-being than their non-envatted counterpart with authentic mental states with identical intrinsic properties. Veridical mental state theories are therefore better placed to capture our intuitions about well-being in these types of cases.

Conditional mental state theories can solve the problem of inauthentic mental states. However, these views also come with a cost. They abandon the conceptual parsimoniousness that is often seen as an attractive feature of mental state theories, and as a consequence they subject themselves to accusations of arbitrariness. This is particularly apparent when proponents of these theories are pressed to stipulate exactly in which way and to what extent the inauthenticity of a mental state affects well-being. The question of how much deception and misperception detracts from well-being does not have an obviously non-arbitrary answer. Standard mental state theories face no such explanatory burden. The idea that if something is non-instrumentally good for well-being, then it is non-instrumentally good in all its instances has intuitive pull. Nevertheless, this may be a worthwhile trade-off if conditional mental state theories can better capture key intuitions about what improves and diminishes well-being.

While conditional mental state theories can answer the problem of inauthentic mental states, I think that there are good reasons to reject these theories. This is because they incur significant costs in addressing this problem, while failing to capture connected intuitions about the importance of objective reality to well-being. There are no mental state theories that can capture these connected intuitions. Consequently, we are better off adopting a non-mental state theory of well-being if we are moved by these intuitions.

Veridical mental state theories are motivated by the desire to capture the importance of objective reality in our theory of well-being. These views do so by making certain mental states a necessary, but not sufficient, condition for improving well-being. Yet there is far more to the intuition that objective reality matters to well-being than simply discounting the effects of inauthentic mental states on well-being. Consequently, conditional mental state theories only partially address the intuitions that motivate them.

Many people have strong intuitions about the importance of objective reality to well-being. While conditional mental state theories can discount the well-being generated from inauthentic mental states, they cannot claim that well-being is affected by states of affairs

that do not produce any mental states whatsoever. If there are states of affairs that produce these effects, then conditional mental state theories are wrong. Bernard Williams has some useful remarks that help illustrate the force of intuitions about the relevance of objective reality to well-being. He points out that sometimes we desire things to happen independently of desiring to gain beliefs or perceptions about whether they have happened (Williams 1973, 261). Without beliefs or perceptions about whether something has happened, no mental states can be generated in response to its occurrence.

Williams distinguishes between three separate types of desire that are often held simultaneously; these are 'my wanting that P, my wanting to have some beliefs on the subject of whether P, and my wanting my beliefs to be true' (1973, 262). However, he notes that there are cases where the three do not coexist. It is possible that we may want a state of affairs to happen but that we may not want to gain knowledge about the occurrence of that state of affairs. For example, someone may desire their former partner to be happy and find love again, while also desiring to not gain knowledge of that love for fear of the emotional turmoil that would accompany such knowledge.

Derek Parfit applies this type of observation to philosophy of well-being and argues that, in at least some cases, our well-being intuitively seems to be affected by objective states of affairs irrespective of any mental states generated in response to them. He points to the case of parenthood and argues that 'even if I never know it is bad for me both if I am deceived and if I turn out to be an unsuccessful parent' (Parfit 1984, 495). The case of the parent-child relationship may be one where the parent's desire for the child's life to be good is often vastly more important to them than their desire to have true beliefs about how well their child's life is going. If Parfit is right, then there are instances where the importance of objective reality matters irrespective of what mental states our perceptions and beliefs produce or fail to produce.

For example, a parent whose child seeks a new life far away in a time before telecommunications would sometimes not receive further information about their child. Presumably parents in this situation would often retain long-standing desires that the absent child lives a good life. Due to the lack of information, no mental states can be generated in response to whether the child lives a good life. In this situation, it seems intuitive to claim that the objective situation of the child affects the well-being of the parent irrespective of any mental states generated. Conditional mental state theories cannot accommodate the intuition that well-being is improved in such cases because they can only modify the well-being effects of the mental states that they take to determine well-being.

It is possible to deny that these types of examples leverage enough intuitive force to necessitate the abandonment of conditional mental state theories. However, if this is the approach that an objector takes, then it is worth considering whether there is enough intuitive force for them to warrant complicating the mental state theory by making it conditional in the first place. The move to a conditional mental state theory both complicates the view and fails to capture the range of intuitions that dissenters have about mental state theories in general. For this reason, it seems more compelling to stick to a standard mental state theory if one is unmoved by intuitions about the importance of objective reality to well-being. If one is moved by these intuitions, then it is best to look for a non-mental state theory that can capture them instead. Desire theories of well-being are good candidates for doing so.

I have argued that conditional mental state theories cannot address the depth of intuitions about the importance of objective reality that motivate them. Moreover, these theories lose considerable intuitive appeal by sacrificing the conceptual parsimoniousness of standard mental state theories. This makes them far more vulnerable to accusations of arbitrariness. In return for these costs, they can avoid the problem of inauthentic mental states. However, they are unable to claim that events that do not produce any mental states whatsoever can affect well-being. Williams and Parfit illustrate why we may sometimes want to claim that such events can affect well-being. Desire theories of well-being face no such problems of valuing objective reality. They can claim that the fulfilment of a desire improves well-being irrespective of any mental states that the fulfilment generates. Consequently, these theories capture intuitions about well-being that mental state theories cannot. Therefore, desire theories of well-being are not made redundant by conditional mental state theories.

Nevertheless, hedonism and broader experientialist theories do get something right about well-being. It strikes me as undeniable that certain mental states always non-instrumentally improve well-being, while others always non-instrumentally diminish well-being. To my mind, a theory of well-being is implausible if it fails to capture the intuition that severe psychological pain always diminishes well-being. Of course, the benefits and harms of certain mental states are perfectly capable of being outweighed by countervailing benefits and harms. Nevertheless, there seems to be at least something good about euphoria and something bad about anguish that is true of every instantiation of these feelings. Desire theories of well-being can capture this intuition by postulating standing desires and aversions to certain mental states (Baber 2010, 252). On this view, human psychology is structured in such a way that we always desire certain mental states and are averse to others. This explains how the acquisition of some mental states is always relevant to well-being. Consequently,

desire theories of well-being can capture the intuition that undergirds hedonist and broader experientialist theories of well-being. The idea that we have standing desires and aversions of this sort plays an important role in my discussions in §3.4; §4.3 and §5.4.[17]

1.4 Against objective list theories

I have outlined some of the attractions of desire theories of well-being and some of the reasons to prefer them to mental state theories. However, there is another family of views that is worth considering. These views fall under the umbrella of objective list theories. This is an extremely heterogeneous group of theories (Woodard 2013, 790). Consequently, it is challenging to give a general treatment of them. Rather than attempting this unrewarding task, I limit myself to a more modest set of objectives in this section. I begin by putting forward a working definition of objective list theories suitable for my purposes. I then outline the Aristotelian position, which is probably the most influential tradition within this family. I end by arguing that even versions of the objective list theory that capture some of the attractions of desire theories of well-being do not eclipse the appeal of these theories.

Objective list theories are united by the claim that well-being is determined by a list of goods and bads. These goods may include things such as friendship, knowledge, and the capacity to exercise power over one's environment. According to the way that these theories are normally formulated, the acquisition of the things on this list is what determines well-being. However, this is not particularly illuminating. The definition is broad enough to encompass almost every theory of well-being. For instance, according to this definition, hedonism can also be conceived of as an objective list of goods and bads (Fletcher 2013, 209–210). It just happens to be a very short list. Consequently, the definition of objective list theories needs refinement. Rather than becoming mired in taxonomic debates, I propose that we conceive of objective list theories as views that claim that there is a list of well-being goods and bads that is not entirely exhausted by mental states and states of pro-attitude fulfilment and frustration. It is sufficient for my purposes here to provisionally define objective list theories as a residual category that encompasses every view not exhausted by mental state theories and pro-attitude theories of well-being (Woodard 2013, 790). This definition captures the views discussed in this section. Whether this is a compelling definition of what unites these

[17] I am using the term 'aversion' throughout this thesis to simply denote a desire for something not to happen.

theories is not my concern here. Consequently, readers are free to substitute this definition with an alternative providing that it also captures these views.

Probably the most well-known objective list theory originates in Aristotle's writings. The central discussion running through his *Nicomachean Ethics* is about what determines whether a person lives a 'good life' (1095a18–21). He uses the term 'eudaimonia' to express this concept. Eudaimonia has two distinct components. The first can be expressed as being the 'good for a person'. This is the same concept as well-being. The second tracks the quality of what makes a 'good person'. This is what determines how good a person is at conforming to the normative standards that govern ideal human behaviour. For the purposes of this discussion, we can simplify this somewhat as simply being about the norms that govern good ethical behaviour. It has been pointed out that intuitively the criteria for ethics and well-being appear to diverge and conflict (Wilkes 1978, 556). This suggests that Aristotle is eliding two separate values under a single term. Nevertheless, this appears to be the view that he and many of his contemporaries held. In their defence, the popularity of the distinction between well-being and ethical value is a relatively modern phenomenon. A compensatory advantage of Aristotle's approach is that theories that account for both values have additional explanatory worth.

Central to Aristotle's conception of eudaimonia is the contemplation of knowledge (1177a18). He also outlined a list of political and ethical virtues, the exercise of which he considered to lead to a secondary type of eudaimonia (1178a9–10).[18] Aristotle derived these standards from what he considered to be the distinct features of human functioning (1097b33–1098a2). Arguably, his reflections on the unique functioning of the gods also play a role in his valorisation of the contemplation of knowledge over the political and ethical virtues (1178b20–25). He explains cases whereby virtuous people are nevertheless immiserated by stipulating a list of instrumental and external goods that are necessary conditions for achieving eudaimonia. He takes these instrumental goods to be necessary for virtuous behaviour. The examples he gives are of relaxation (1176b34–1177a1), leisure (1177b4–5), and friendship (1169b17–20). The absence of these goods impedes our ability to exercise the virtues necessary for eudaimonia. Conversely, Aristotle thought that external

[18] There is a debate among scholars of Aristotle between 'inclusive' and 'dominant' readings of his view of eudaimonia (Hardie 1965, 279). According to the inclusive view, eudaimonia requires a mixture of contemplation, and the political and ethical virtues. Conversely, the dominant view takes Aristotle's statement that 'happiness extends, then, just so far as contemplation does' (1178b29) to override his discussion of the importance of the ethical and political virtues for eudaimonia. I set aside this debate here.

goods were not necessary to behave virtuously. However, without them, the exercise of virtue does not lead to eudaimonia (Heinaman 1993, 35). Some of Aristotle's examples of external goods include good social and familial connections (1097b9–11), and the absence of catastrophic events in life (1100b10–13). This allows him to explain how classical bastions of virtue, such as Priam in Homer's *Iliad,* nevertheless do not achieve eudaimonia (1101a5–10).

Aristotle is clearly wrong about the specifics of what determines well-being. It is implausible that the contemplation of knowledge is the primary way of improving well-being. However, there is something compelling about the methodology that he uses for identifying those specifics. Perhaps the standards of human well-being are derived from facts about the sorts of creatures that we are. This method at least gives us a non-arbitrary way of identifying species-specific[19] standards of well-being. Moreover, grounding well-being in our nature may serve as a good basis for postulating a plurality of distinctly human goods. Perhaps we ought not to expect complex creatures such as us to have a monistic standard of well-being. Perhaps the monistic theories of well-being purchase their conceptual parsimoniousness at the expense of oversimplifying their explanations about the type of creatures that we are. However, while the form of the Aristotelian theory is attractive, problems emerge the moment that we begin to fill in the specifics. This is because proposed goods often fail to resonate with the range of subjects to which they are supposed to apply. For instance, education appears to be a good candidate for being an objective good for humans. Nevertheless, for some people education does not resonate with any of their pro-attitudes. A theory that postulates that education is an objective good means that its attainment improves well-being even if it does not resonate with any of a subject's pro-attitudes. This is counterintuitive. This argument is repeatable with many other candidates that objective list theories might put forward as well-being goods. If objective list theories postulate alienated standards of well-being, then it seems there is good reason to reject these theories.[20]

---

[19] 'Species-specific' need not refer to members of a biological kind. It could be that differences in capacities determine the type of species membership that is relevant to well-being (Yelle 2016, 1415). For instance, some great apes and most adult humans may belong to the same well-being species if they share certain capacities, while human babies who lack these capacities may belong to different well-being species.

[20] In defence of alienation, it can be pointed out that the fulfilment of some desires appears to lessen well-being. This is often said to be the case of a subsection of desires that emerge under conditions of oppression or limitation (Khader 2011; Sen 1984; Elster 1983). People with such adaptive desires may be alienated from the things that would actually make their lives go better for them. If this is right, then perhaps we ought not to expect someone's own well-being to always resonate with them.

Eden Lin has proposed a hybrid-theory of well-being (2017a, 372). This view captures the resonance requirement. It does so by postulating a list of objective goods, the attainment of which only improves well-being on the condition that these goods resonate with the subject who attains them. Thus, to take my earlier example, friendship, knowledge, and the capacity to exercise power over one's environment only improve well-being if these goods resonate with the subject who acquires them. A theory of this type avoids the objection that objective list theories are committed to an alienated standard of well-being.

Guy Fletcher outlines a different way in which objective list theories can capture the resonance requirement. He points out that these theories can postulate a list of goods that contain pro-attitudes as necessary components (Fletcher 2013, 216). He suggests that the following goods have this structure: achievement, friendship, happiness, pleasure, self-respect, and virtue (Fletcher 2013, 214). These goods involve a subjective and an objective component. A precedent for this approach can be found in Susan Wolf's definition of meaningfulness. According to her, 'meaning arises when subjective attraction meets objective attractiveness' (Wolf 1997, 211). This definition of meaning also requires both a subjective pro-attitude and an objective value to which that attitude is directed towards. Regardless of whether we agree with the specifics, the general approach is a distinct way that objective list theories can avoid alienation. Consequently, some objective list theories are well-placed to capture the resonance requirement.

Nevertheless, problems persist. Firstly, there is no obvious way to ascertain the relative importance of the items of the objective list (Fletcher 2013, 214). Desire theories of well-being and monistic mental state theories can appeal to the intensity of the relevant desires or attitudes to determine the extent to which their fulfilment or acquisition improves well-being. Objective list theories have no obvious way of working this out. Secondly, objective list theories must complicate themselves further if they want to account for well-being's negative dimension. When doing so it seems likely that they will need to postulate asymmetric goods and bads (Kagan 2014, 272–284). This is because there appear to be goods that do not have corresponding bads and vice versa. For instance, friendship is a good candidate for an objective good, but enmity is an unconvincing candidate for a corresponding bad. Desire theories of well-being and many mental state theories do not suffer from this explanatory burden. Instead, they postulate a pleasing symmetry between what improves and what decreases well-being. Finally, objective list theories face the problem of how to identify what goods and bads belong on the list. The Aristotelian appeal to human nature sounds plausible until it is invoked to support a specific list of goods and bads. Until progress is

made on this problem, no specific version of the objective list theory is likely to win much popular support. While some versions of the objective list theory can capture the resonance requirement, the explanatory burden that these theories face means that they do not eclipse the appeal of desire theories of well-being.

In this section I have focussed on discussing those objective list theories that capture one of the key attractions of desire theories of well-being. This is the resonance requirement. I have argued that, while some objective list theories are well-placed to capture this requirement, they nevertheless remain beset by a series of problems that undermine their attractiveness. Most notably, these theories face the challenge of how to non-arbitrarily populate the objective list of well-being goods and bads. This does not show that these theories are wrong. Nevertheless, their inability to eclipse the appeal of desire theories of well-being goes some way towards illustrating why extended discussion of these latter theories is warranted.

1.5 Chapter summary

This chapter has set up the preliminaries of my thesis. I have explained what well-being is, what assumptions I am making about it, and the methodology I am using to investigate it (§1.1) I then outlined desire theories of well-being, some attractions of these views, and how these attractions apply to popular versions of the theory (§1.2). Afterwards I argued that these theories are not made redundant by versions of the mental state theory that attempt to capture aspects of their appeal (§1.3). I then extended a similar treatment to objective list theories that capture the resonance requirement (§1.4). This lays the ground for my investigation into the problems that threaten to undermine desire theories of well-being. In my next chapter, I will turn to the first of these problems. This problem claims that desire theories of well-being are incorrect because they entail that self-sacrifice is impossible.

Chapter 2 The Problem of Self-sacrifice

2.0 Introduction

This chapter considers one of the most influential objections to desire theories of well-being. Mark Overvold argues that these theories entail that self-sacrifice is impossible. If this is correct, then desire theories of well-being are undermined by the counterintuitive nature of their implications. I argue that a new solution to this problem can be found by rejecting proportionalism about desire and motivation. This is the view that desires always motivate proportionally to their strength. Rejecting this view allows us to account for problem cases of self-sacrifice by claiming that they occur when our altruistic desires motivate us with disproportional strength. This approach solves an influential problem for desire theories of well-being and enriches our understanding of the psychology of self-sacrifice.

This chapter has the following structure: §2.1 outlines the problem of self-sacrifice for desire theories of well-being. §2.2 surveys existent proposed solutions to the problem and highlights their shortcomings. §2.3 puts forward a new approach to this problem based on the rejection of proportionalism about desire and motivation. §2.4 highlights independent reasons to reject proportionalism about desire and motivation. §2.5 concludes with a summary.

2.1 The problem of self-sacrifice

It is often said that viable theories of well-being must have intuitive implications about paradigm cases of benefit and harm (Fletcher 2016, 10). Mark Overvold claims that desire theories of well-being fail in this respect (1980). He argues that this is because these theories entail that self-sacrifice is impossible. According to Overvold, for an action to qualify as self-sacrificial it must contain three features. Firstly, it must result in an anticipated loss of well-being for the person who performs it. Secondly, it must be voluntary. Thirdly, there must be a viable alternative action that the agent correctly predicts would be more in their self-interest to perform (Overvold 1980, 109–114).

Overvold's model of self-sacrifice is incomplete. He underspecifies the time at which an action must diminish well-being in order to qualify as self-sacrificial (Rosati 2009, 315). It is unclear whether Overvold thinks that the loss of well-being must happen within the limited

aftermath of an action, or whether self-sacrifice must result in a net loss of lifetime well-being. If Overvold has in mind the former specification, then his view entails that sometimes self-sacrifice improves lifetime well-being. This is because things that diminish well-being in the short-term may nevertheless result in net improvements to lifetime well-being. If he has in mind the latter specification, then this view entails that we cannot definitively classify an action as self-sacrificial until it has ceased to have ramifications on well-being. This is because the consequences of an action can oscillate over time between being a net loss and a net improvement to a subject's lifetime well-being.

If we want to preserve Overvold's model of self-sacrifice, then we must either accept that self-sacrifice can improve net lifetime well-being, or reserve judgement about whether actions are self-sacrificial until after their causal chain of consequences has ended. It strikes me that neither position is particularly counterintuitive. If we embrace the first, then we can claim that actions that result in a net improvement to lifetime well-being can qualify as self-sacrificial providing that these improvements are unforeseen incidental consequences that occur after the immediate consequences of the action have ceased. Alternatively, if we embrace the second, then we can accept that while we cannot know whether an action is self-sacrificial until after its causal chain of consequences has ended, we can nevertheless describe actions as altruistic, virtuous, or provisionally self-sacrificial in the interim. I leave readers to decide between which specification they find more convincing.

In what follows I assume that Overvold's position is broadly correct.[21] However, not much hinges on this. Alternative formulations of self-sacrifice raise similar issues for desire theories of well-being, and other writers have put forward related problems.[22] Responses to Overvold can be adapted and applied to these problems.

---

[21] Overvold's view is contestable. For instance, we may want to claim that actions must be supererogatory in order to qualify as self-sacrificial (Smilansky 2021, 2). We may also want to soften or reject Overvold's first condition if we think that non-optimific actions that involve forgoing significant increases in well-being can qualify as self-sacrificial. Furthermore, we may want to add an additional requirement that actions must aim at something good or valuable in order to be self-sacrificial (Rosati 2009, 314–315). Additionally, we may think that the prospective consequences, whether objective or subjective, of an action, rather than its actual consequences, are what makes an action self-sacrificial (Archer 2016, 338–339). Finally, we may want to exclude trivial losses of well-being from qualifying an action as self-sacrificial (Rosati 2009, 316). There are undoubtedly further debates to be had over the definition of self-sacrifice. However, I set them aside here.

[22] Chris Heathwood has compiled a list of related arguments by luminaries including Richard Brandt, James Griffin, Thomas Carson, Thomas Schwartz, L. W. Summer, and Amartya Sen (Heathwood 2011, 18–19).

Overvold argues that if desire theories of well-being are correct, then informed voluntary actions can never contain the first or third features of self-sacrifice. His view is that we cannot anticipate losses to our well-being from our informed voluntary actions because they always improve our well-being. Moreover, he claims that there can never be alternative actions that we anticipate are more in our self-interest to perform. This is presumably because he takes our voluntary actions to always be produced by our presently strongest balance of desires. Our presently strongest balance of desires is directed towards the action that we predict will maximise the fulfilment over frustration of our presently existing desires, when adjusted for the relative strength of the desires in question. Consequently, our informed voluntary actions are always optimal for our well-being (Overvold 1980, 115). If an action is involuntary, then it can fulfil the first and third criteria of self-sacrifice. However, by definition, it falls foul of the second.

Self-sacrifice is a paradigm case of harm that viable theories of well-being must capture. To illustrate this point, Overvold constructs an example of a father who desires to kill himself for the life insurance payout to fund his children's education (1980, 107). He takes this to be an uncontroversial case of self-sacrifice. However, desire theories of well-being seem to counterintuitively entail that the father's suicide would not diminish his well-being. Stephen Darwall provides an alternative example that is not complicated by the additional questions about the harm of death that Overvold's case raises. He asks us to imagine a woman called Sheila who deeply desires to financially contribute to the reconstruction of a war-ravaged city (Darwall 2002, 43–44). Sheila has a degenerative disease that, if left untreated, will leave her with severe memory loss and confused about where her money has gone. Fortunately for Sheila, her disease is curable with relatively inexpensive medication that has no side effects. Unfortunately for her, she prefers to spend the money on the reconstruction efforts rather than the medication. Were Sheila to act upon this preference, then intuitively she would have acted self-sacrificially. Yet, as with Overvold's example, desire theories of well-being seem to counterintuitively entail that Sheila's action improves her well-being instead. If this is correct, then we have good reason to reject these theories. I will refer to this argument as 'the problem of self-sacrifice'.

The problem of self-sacrifice can be summarised as follows:

P1: Self-sacrifice is a harm that viable theories of well-being must capture.

P2: Desire theories of well-being fail to capture the harm of self-sacrifice.

C: Therefore, desire theories of well-being are not viable theories of well-being.

Overvold's argument is valid. Moreover, premise one strikes me as unassailable. However, premise two is contestable. If it can be shown to be incorrect, then an influential objection to desire theories of well-being will have been defused.

2.2 Proposed solutions to the problem of self-sacrifice

Premise two of the problem of self-sacrifice rests upon at least five unexamined assumptions. These are:

1. Desire theories of well-being claim that the fulfilment or frustration of any desire non-instrumentally affects well-being.[23]

2. Self-sacrificial actions necessarily reduce the well-being of the person who performs them.

3. Human psychology is structured in such a way that altruistic actions are necessarily motivated by desires.

4. Desire theories of well-being claim that maximising present desire satisfaction is always optimal for our well-being.

5. Human psychology is structured in such a way that the strength of our motivations is always proportional to the strength of our desires.

Overvold takes assumption one to be part of the definition of desire theories of well-being, and assumption two to be part of his definition of self-sacrifice. Assumptions three, four and five are presupposed but not explicitly articulated in his paper (Overvold 1980). Proposed solutions to the problem of self-sacrifice can be constructed from the rejection of any one of Overvold's assumptions. I consider them sequentially in the following subsections.

2.2.1 Do all desires count?

Overvold's first assumption concerns the definition of desire theories of well-being. He takes these views to claim that the fulfilment or frustration of any desire non-instrumentally affects

[23] Following Richard Brandt (1972, 682), Overvold restricts the desires that determine well-being to those that are fully informed (1980, 107). Nevertheless, his problem emerges regardless of whether we accept this restriction. For simplicity, I omit further mention of it.

well-being. However, this is not true of all versions of the theory. For instance, some writers argue that these theories ought to restrict which desires non-instrumentally affect well-being to only our self-regarding desires (Carson 2000, 75; Parfit 1984, 494; Sidgwick 1907, 109–110). These are the desires that we have for states of affairs that involve our own lives in some way. If this theory is right, and the desires that motivate self-sacrifice are non-self-regarding, then the fulfilment of these desires does not non-instrumentally improve well-being. Consequently, actions motivated by non-self-regarding desires can qualify as self-sacrificial according to Overvold's criteria.[24] It seems likely that the prospective actions described in Overvold's example of the father and Darwall's example of Sheila fall into this category.

However, as many writers have pointed out, it is not obvious how to non-arbitrarily distinguish between self- and non-self- regarding desires (Portmore 2007, 27; Adams 1999, 88; Sobel 1997, 506). One approach claims that desires can be identified by their content in some way. For instance, it may be that self-regarding desires have an 'I' in their content.[25] However, this view is subject to counterexamples. For example, if I desire that 'I help you attain your goal of learning to ride a bike', then this intuitively appears to have both self-regarding (that I help you) and non-self-regarding (that you attain your goal) components. However, because this desire has an 'I' in its content, this approach relegates my desire to simply being self-regarding. Nevertheless, if I fulfil my desire to help you learn to ride a bike at considerable cost to myself, for example, by buying you the bike, then it strikes me as counterintuitive to claim that the fulfilment of this desire contributes to my well-being as much as that of any other equally strong self-regarding desire. Desires to be of service to others are counterexamples to this way of distinguishing between self-regarding and non-self-regarding desires (Adams 1999, 140).

Moreover, this approach entails that fulfilling our desires to live up to our duties and obligations necessarily benefits us (Darwall 2002, 30). It also entails that fulfilling the

[24] Overvold suggests his own restriction for how desire theories of well-being can handle the problem of self-sacrifice. His proposal is that only our desires that logically require our own existence as a constituent part count towards our well-being (Overvold 1980, 117–118 n.10). This proposal is more restrictive than the one discussed herein and is susceptible to a similar critique.

[25] This approach is premised on the idea that desires are, or can be accurately represented as, propositional attitudes (Sinhababu 2015; McDaniel & Bradley 2008, 268). Some writers have argued that this is incorrect (Brewer 2006; Thagard 2006). If they are right, then we cannot distinguish between desires in this way. Nevertheless, in this thesis I assume that the claim that desires are, or can be accurately represented as, propositional attitudes is broadly correct. I make arguments in §3.3, §5.3 and §5.4 that depend upon this view.

desires that govern our own agent-centred moral injunctions improve our well-being (Sobel 1998, 267). This is because these desires have the person who holds them in their content. Nevertheless, fulfilling desires of this type does not always seem to improve our well-being. For instance, fulfilling the desire to adhere to the moral injunction against committing tax fraud does not always seem to be in everyone's self-interest. Indeed, the anticipated prudential benefit to oneself of violating this injunction is largely what motivates people to do so. The prudential benefit to oneself that comes at the expense of the common good is part of what makes this type of action reprehensible. However, this version of the restricted desire theory of well-being seems to classify fulfilling the desire to adhere to this moral injunction as non-instrumentally improving our well-being. Consequently, while restricting which desires count towards well-being in this way makes self-sacrifice possible, it does so at the expense of leading to counterintuitive implications about which things improve well-being.

Furthermore, while it is counterintuitive to conclude that self-sacrifice cannot exist, it also seems wrong to claim that many of our altruistic actions do not non-instrumentally improve our well-being. Yet, if this restriction is correct, then altruistic desires which do not have the agent who holds them in their content do not non-instrumentally improve that agent's well-being when fulfilled. This is a necessary commitment of excluding non-self-regarding desires from non-instrumentally affecting well-being. However, common intuitions suggest that many altruistic actions do improve the well-being of the altruist (Wolf 1982, 437). This version of the restricted desire theory of well-being fails to capture these intuitions. Moreover, it also counterintuitively excludes other types of non-self-regarding desires from non-instrumentally affecting well-being. For example, it excludes our desires about our children and our desires about the fortunes of our favoured sports teams from doing so (Heathwood 2011, 25). While it may be possible to tinker with this way of distinguishing between self- and non-self-regarding desires to address some of these shortcomings, it strikes me that any view that relies upon such a distinction is destined to make casualties of our intuitions in at least some of these cases. Consequently, we are better off looking elsewhere for a solution to the problem of self-sacrifice.

More recently, Chris Heathwood has put forward a version of the desire theory of well-being based upon a popular distinction between two senses of the term desire (Heathwood 2019; Schueler 1995, 1; Vadas 1984, 273). Though Heathwood does not himself rely upon this distinction to solve the problem of self-sacrifice (§2.3.3), it might be thought that the distinction can help with this problem. According to the behavioural sense, desires are

simply any psychological state that motivates us. Conversely, according to the genuine attraction sense, desires are only those motivational states that are characterised by enthusiasm, appeal, interest, excitement, and attraction (Heathwood 2019, 673). Heathwood argues that desire theories of well-being ought to restrict which desires affect well-being to only the latter sense. If we adopt this approach, then we can explain why acts of self-sacrifice do not non-instrumentally improve well-being. This is because these acts are motivated solely by desire in the behavioural sense.

There are at least three reasons to be sceptical of this solution. Firstly, it is unclear whether the distinction that undergirds it holds up. It is possible to have a desire that begins by encompassing the features of genuine attraction, progresses onto having purely behavioural features, before returning to having the features of genuine attraction. If the two senses of desire are distinct, then it is surprising that desires can oscillate between them while retaining the same content. The fact that desires can fluctuate in this way suggests that we are dealing with one sense of desire that has both motivational and phenomenological components. This point is reinforced by the observation that both the genuine attraction and behavioural senses of desire come in degrees. For instance, people frequently experience only moderate motivation towards outcomes that they have a strong phenomenological attraction towards. The existence of gradations in these effects suggests that desire is not divisible in this way.

Secondly, even if the distinction holds up, it is unclear whether well-being ought to be based upon it. Doing so leads to some counterintuitive implications. People experiencing anhedonia may, for instance, not feel much enthusiasm or excitement about pursuing their desires (Tully 2017, 4). Nevertheless, sometimes the behavioural aspects of their desires are left unaffected. We should be reticent to accept, at least in mild cases, that people in this condition do not have many desires relevant to well-being. After all, if they still experience desire in the behavioural sense, and previously did experience it in the genuine attraction sense, then it is counterintuitive to claim that the fulfilment or frustration of these desires never non-instrumentally affects their well-being. Anhedonia is not alone in suppressing the genuine attraction sense of desire, while leaving the behavioural sense unaffected. Exhaustion typically also weakens our enthusiasm for pursuing our desires. It is similarly counterintuitive to fully discount the well-being generated from fulfilling desires that we no longer experience genuine attraction towards due to exhaustion. Consequently, there are good reasons to think that well-being ought not to be based upon this distinction.

Thirdly, even if the distinction holds up and is attractive to base well-being upon, this approach fails to satisfactorily solve the problem of self-sacrifice. This is because not all

cases of self-sacrifice are motivated purely by the behavioural sense of desire. For example, consider the parent who sacrifices themselves to prevent terrible harm to their children. When faced with the consequences of inaction, they may find that their prospective sacrifice incites enthusiasm in them. Heathwood's restriction fails to capture the intuition that such acts qualify as self-sacrificial. Therefore, we ought to look elsewhere for a solution to the problem of self-sacrifice.

2.2.2 Does self-sacrifice always diminish well-being?

The problem of self-sacrifice can be solved if we reject Overvold's second assumption. This is the view that self-sacrifice necessarily involves a loss of well-being. Instead, it may be that the 'sacrifice' in self-sacrifice refers to the voluntary loss of something else. If this is right, then there need be no incompatibility between self-sacrifice and desire theories of well-being. This is the approach that Connie Rosati takes (2009). She argues that although self-sacrifice often involves a loss of well-being, such a loss is neither a necessary nor sufficient condition for actions to qualify as self-sacrificial. Instead, she argues that self-sacrificial actions must involve the loss of part of the self. Rosati points out that paradigm cases of self-sacrifice tend to involve such losses (2009, 317). For instance, sacrifices of life and limb seem to best capture our intuitions about what self-sacrifice paradigmatically looks like. An advantage of defining self-sacrifice in this way is that it explains what makes these cases paradigmatic.

Rosati's argument relies upon a distinction between parts of the self and more peripheral aspects of our lives. She presents a complex picture of the self that involves a collection of different constituents (Rosati 2009, 316–319). These include physical, emotional, and psychological integrity. Some of our projects, activities, relationships, and interests in life are also constitutive of the self. If Rosati is right, then desire theories of well-being are compatible with the existence of self-sacrifice. This is because actions can qualify as self-sacrificial despite advancing the well-being of the person who performs them (Rosati 2009, 314). Consequently, this view can explain why the father in Overvold's example would be acting self-sacrificially if he kills himself to further the educational opportunities of his children. This is because he would be sacrificing his self. Similarly, this view can also explain Darwall's example of Sheila forgoing life-changing medication in order to further the reconstruction efforts of a war-torn city as a case of self-sacrifice. This is because Sheila

sacrifices her psychological integrity, which is a constituent part of her self. Therefore, Rosati's argument can solve the problem of self-sacrifice.

Nevertheless, there are several reasons to be sceptical of Rosati's proposed solution to this problem. Firstly, Rosati's distinction between sacrifices of the self and sacrifices of well-being rests upon a particular conception of what constitutes the self. This conception is underspecified in her paper (Rosati 2009). Rosati's picture of the self does not provide clear guidance on the status of specific projects, activities, relationships, and interests in life. It is unclear what conditions these things must satisfy in order to qualify as parts of the self. Without further specification, this invites indeterminacy about which actions are self-sacrificial. Moreover, it is unclear what unifies the collection of things that Rosati lists as constituents of the self. Without a unificatory explanation, Rosati's definition postulates many disparate constituents. This risks being intolerably arbitrary. Consequently, this approach is a conceptually costly solution to the problem of self-sacrifice. We should be reticent to incur this cost if there are more economical alternatives available.

Secondly, even if we can construct a satisfactory theory of the self, it is counterintuitive to base our definition of self-sacrifice upon it. Rosati's view entails that even significant losses to well-being have no bearing on whether an action is self-sacrificial. Thus, if you were to donate all your savings to charity and forgo the luxuries that they would have afforded, then this only counts as self-sacrifice if the donation results in the loss of sufficiently important projects, activities, relationships, and interests in life. Or, if the donation leads to a loss of physical, emotional, or psychological integrity. It is possible to imagine circumstances where this need not be the case. Nevertheless, the significant loss of well-being that most of us would incur as a consequence of donating all of our savings to charity suggests that doing so generally ought to be classified as an act of self-sacrifice. It would be revisionary of our intuitions about self-sacrifice were our theory to fail to classify this type of action as self-sacrificial.

Thirdly, Rosati's account entails that actions that significantly advance our well-being at the expense of compromising a comparatively minor aspect of our self counterintuitively qualify as self-sacrificial. For instance, imagine that you compromise your physical integrity by donating a kidney to a stranger. You do this with altruistic intentions and predict no compensatory benefits from your action. Nevertheless, immediately after the procedure you wake up to find out that the beneficiary of your kidney is a billionaire who rewards you with a significant share of their fortune. You use this money in such a way that significantly improves your well-being in both the immediate aftermath of the operation and over the

course of your life. In this case, you have sacrificed a comparatively minor aspect of your self, while benefitting from a large increase in both immediate and lifetime well-being. However, Rosati's view classifies this action as self-sacrificial. This is counterintuitive. While this action is praiseworthy and undertaken with a self-sacrificial attitude, it strikes me as ill-described as a self-sacrificial action. Examples like this illustrate that our notion of self-sacrifice is inextricably entwined with the concept of well-being.

Finally, Rosati's argument does not explain how desire theories of well-being can account for cases whereby we knowingly act in such a way that diminishes our own well-being. Even if we accept Rosati's definition of self-sacrifice, it is a problem for desire theories of well-being if they cannot account for such cases. Consequently, Rosati's argument takes us no closer to an intuitive understanding of Overvold's example of the father whose desires lead him to sacrifice his life for the benefit of his children. Nor does it take us closer to an intuitive understanding of Darwall's example of Sheila whose desires motivate her to forego life-changing medication in order to further the reconstruction efforts of a war-ravaged city. In both cases, we should be able to recognise the substantial loss of well-being that these people voluntarily incur. While Rosati's framework allows us to call actions self-sacrificial, it does not allow us to account for voluntary informed actions that result in a loss of well-being. Consequently, we should reject Rosati's proposed solution to the problem of self-sacrifice.

2.2.3 Do only desires motivate?

Another approach to the problem of self-sacrifice involves rejecting Overvold's third assumption. This is the view that altruistic actions are always motivated by desires. The view that voluntary actions are necessarily motivated by desires is most often referred to as the Humean Theory of Motivation. Although Overvold does not explicitly subscribe to this position, his argument is nevertheless premised upon a Humean explanation of altruistic action. This is because desire theories of well-being only entail that altruistic actions non-instrumentally improve our well-being if they fulfil our desires.

If we reject the Humean position, then we can claim that some altruistic actions are not motivated by desires. These actions may instead be motivated by alternative mental states, such as moral beliefs (Shafer-Landau 2003, 122; Dancy 1993). If this is right, then these

actions do not necessarily fulfil desires.[26] There is something to be said for the idea that moral beliefs can motivate. After all, we tend to think of moral judgements as the sort of thing that can be correct or incorrect. This is a feature more often ascribed to beliefs than desires. And yet it seems rare that someone is entirely unmotivated by their sincere moral judgements. Consequently, there are some independent reasons to think that moral judgements are beliefs that motivate.[27]

Adopting a non-Humean explanation of altruistic action can harmoniously combine desire theories of well-being with the existence of self-sacrifice. However, this approach involves weighty theoretical commitments. The Humean Theory of Motivation is popular (Sinhababu 2017; Smith 1994). Rejecting it commits proponents of desire theories of well-being to a controversial position within moral psychology. Moreover, this approach means that the viability of desire theories of well-being is contingent upon whichever theory of motivation happens to be correct (Heathwood 2011, 33). This is precarious grounding for any theory of well-being.

Furthermore, while this argument makes self-sacrifice possible, it nevertheless leaves us in the unenviable position of conceding that we can never sacrifice ourselves if we are motivated solely by desires. This seems wrong. Even if we accept that moral beliefs can motivate, there are still cases of self-sacrifice that are more intuitively explained as motivated by desire. For instance, sacrificing oneself to save the life of a loved one is often more intuitively explained as motivated by desire rather than moral belief. This is especially the case when our moral beliefs conflict with our self-sacrificial action. For example, if we know that our loved one will do great harm to others if saved but we choose to save them anyway, then this seems to be a case where the motivational force of our desire outweighs that of our moral belief. Moreover, the examples of self-sacrifice given by Overvold and Darwall are badly explained as motivated by moral beliefs. Both characters described in these examples seem to be motivated by desires. Therefore, even if we accept that moral beliefs can motivate, self-sacrifice is nevertheless sometimes more intuitively explained as

[26] This argument fails if we hold the view that moral beliefs do motivate but that desire is always somehow involved in this process. For instance, it has been argued that a logical consequence of having a motivating belief is that it produces a desire (Nagel 1970a, 30). If this is correct, then an appeal to the motivational effects of moral beliefs will not solve the problem of self-sacrifice.

[27] The debate between Humeans and non-Humeans is longstanding and somewhat intractable. Nevertheless, the non-Humeans have been successful in showing there are at least some reasons to suspect that the Humean position may be mistaken. It is sufficient for my purposes here to draw attention to these reasons.

motivated by desire. Consequently, it is better to remain neutral on the independent viability of the Humean Theory of Motivation and look for alternative solutions to the problem of self-sacrifice.

2.2.4 Is well-being best served by maximising present desire satisfaction?

Another approach to the problem of self-sacrifice involves rejecting Overvold's fourth assumption. This is the idea that desire theories of well-being claim that maximising the satisfaction of our presently strongest balance of desires is always optimal for our well-being. Chris Heathwood challenges this view. He argues that only a specific type of desire theory of well-being accepts this claim. He calls this view Life Preferentism (Heathwood 2011, 22).[28]

Like other versions of the desire theory of well-being, Life Preferentism claims that well-being is solely determined by the fulfilment and frustration of desires. However, it is distinctive for the additional claim that fulfilling our presently strongest balance of desires is always optimal for our well-being. Consequently, to take Overvold's example, if a father's presently strongest balance of desires leads him to kill himself for the life insurance to fund his children's education, then this is the best outcome available to him. Or, to take Darwall's example, if Sheila's presently strongest balance of desires are best fulfilled by forgoing life-changing medication in favour of spending more of her money on a war-torn city's reconstruction efforts, then this outcome is optimal for her. This view entails that actions motivated by our presently strongest balance of desires can never be self-sacrificial.

Life Preferentism is not an attractive theory. Not only does it have counterintuitive implications for our understanding of self-sacrifice, but it also fails to recognise that sometimes fulfilling our presently strongest balance of desires is not in our best interests (Heathwood 2011, 26). This happens often. For example, my presently strongest balance of desires may lead me to socialise late into the night on a weekday evening. Nevertheless, the

[28] Heathwood cites John Rawls as advancing a similar position to this (Heathwood 2011, 22). The difference being that Rawls requires desires to be adequately informed and rational in order for them to count towards well-being (1971, 417). Recently, Alexander Dietz has suggested we call views that only count existing desires as relevant to well-being 'reactive', and those that count future desires 'proactive' (2023). This is perhaps a more helpful terminology as many people use the term 'desire satisfactionism' to refer to desire theories of well-being more broadly. Nevertheless, I will stick with the terminology inherited from Heathwood, as it is his specific views that are under discussion here.

fatigue I experience the following day may frustrate more and stronger desires than those fulfilled by the late-night revelry. In this case, acting on weaker desires would be better for me. The fact that Life Preferentism fails to capture this intuition means that it should be rejected.

Heathwood argues that more plausible versions of the desire theory of well-being fall under the rubric of Desire Satisfactionism (2011, 24).[29] This view claims that well-being is determined by the total balance of desire fulfilment over frustration, when adjusted for strength, over the course of a life. This captures the intuition that sometimes we are made worse off by fulfilling our presently strongest balance of desires. In such cases, our well-being is improved by the fulfilment of our desires, but this improvement is outweighed by the desires that our action frustrates.[30] If we accept Desire Satisfactionism, then the problem of self-sacrifice is surmountable.

Heathwood discusses two ways in which an action can be self-sacrificial. The first occurs when our action prevents future desire fulfilments. This is how Heathwood explains Overvold's case of the father who sacrifices his life to fund the education of his children (Heathwood 2011, 27). In this case, the bulk of the harm arises from the deprivation of future desire fulfilments that are prevented by his suicide.[31] However, not all cases of self-sacrifice are well explained by an appeal to the deprivation of desire fulfilments. In cases where the subject's life does not end as a consequence of their action, it appears more common that the harm of self-sacrifice arises primarily as a result of frustrated desires. Heathwood writes of such cases that 'it is possible for a person to know, even vividly, that he will desire certain things in the future, and yet fail to be moved in the present to behave in such a way that those future desires will be satisfied' (2011, 28). On this view, sometimes we act self-sacrificially in the knowledge that our action will frustrate some of our desires in the future. This seems

[29] This discussion of Heathwood is based on his paper 'Preferentism and self-sacrifice' (2011). More recently, Heathwood has advanced a separate argument based on a distinction between the behavioural and genuine attraction senses of desire (2019). I discussed why we ought to reject this argument in §2.2.1 of this paper.

[30] Heathwood's position has precedents. Wayne Sumner reflects that sometimes 'satisfying the desire made me *to that extent* better off, but it also frustrated other, more important desires, so that on balance I ended up worse off' (1996, 131). Moreover, Thomas Carson points out that sometimes desire fulfilment 'frustrates *other* desires that are of greater importance to the agent' (2000, 73). Henry Sidgwick also acknowledges that sometimes 'the desired result is accompanied or followed by other effects which when they come excite aversion stronger than the desire for the desired effect' (1907, 110). Nevertheless, the implications of this observation are discussed in most detail by Heathwood.

[31] This approach to the harm of death has substantial philosophical precedent (Feldman 1991; Nagel 1970b).

to be what is happening in Darwall's example of Sheila forgoing preventative medication in order to spend her money on the reconstruction of a war-ravaged city. In this case, the harm of foregoing the medication is incurred at a future time when the effects of her disease are experienced. Consequently, Heathwood's view intuitively explains both examples of self-sacrifice that this discussion opened with.

Nevertheless, while Heathwood's view intuitively explains many cases of self-sacrifice, there remain counterexamples that are ill-described by this psychology. Many acts of self-sacrifice feel as if we are acting against our presently strongest balance of desires at the time of action. Consider volunteering your time in a dull but important awareness-raising leafleting campaign about the pernicious health effects of air pollution, rather than spending the day in the park basking in the sun. This experience feels like we are frustrating a presently stronger balance of desires than we are fulfilling. This phenomenology is unaccounted for by Heathwood's position. Moreover, there is no obvious future time at which to point to frustrated desires or the deprivation of desire fulfilments in order to account for the harm that the leaflet deliverer incurs. Instead, the harm appears to be incurred throughout the duration of the action. Consequently, we need an alternative explanation of how desire theories of well-being can account for such cases.

Due to the unviability of Life Preferentism, we ought to reject Overvold's fourth assumption and accept Heathwood's Desire Satisfactionism. However, without further modification, our resultant theory is committed to a counterintuitive explanation of the psychology that motivates some cases of self-sacrifice. While Heathwood's theory makes self-sacrifice possible, it does so by putting forward an incomplete picture of the psychological structure to which all acts of self-sacrifice must conform. According to this picture, it is only through reference to future desire frustrations or the deprivation of future desire fulfilments that we can account for the harm of self-sacrifice. However, this explanation does not explain all cases of self-sacrifice in an intuitive way. Consequently, we need an additional explanation of how desire theories of well-being can explain residual cases of self-sacrifice that are badly accounted for by Desire Satisfactionism.

2.3 Desire and motivation in desire theories of well-being

An intuitive explanation of self-sacrifice should account for the experiential quality of acting against our presently strongest balance of desires. Recall that our presently strongest balance

of desires is for the action that we predict will maximise the fulfilment over frustration of our presently existing desires, when adjusted for the relative strength of the desires in question. It is directed at the action that we overall most desire to perform. Desire theories of well-being can capture the experiential quality of acting against our presently strongest balance of desires by rejecting Overvold's fifth assumption. This claims that human psychology is structured in such a way that the strength of our motivations is always proportional to the strength of our desires. We can term this view 'proportionalism about desire and motivation' (henceforth proportionalism).[32] A commitment to proportionalism explains Overvold's assumption that we always act upon our presently strongest balance of desires.

If we reject proportionalism, then we can account for cases of self-sacrifice that are ill-described by Heathwood's Desire Satisfactionism. In such cases, our presently strongest balance of desires does not cause our self-sacrificial action. Instead, our action is caused by a weaker balance of desires that motivates us with disproportional strength. Many self-sacrificial actions that are motivated by desires to live up to our obligations or duties are well explained by this approach. In such cases, the phenomenological quality of the action suggests that we are acting against our strongest balance of desires. This seems to be what is happening in cases such as that of the leaflet deliverer. Our feelings of sacrifice arise because, while we do desire to undertake the self-sacrificial action, we have other presently stronger desires that are frustrated by doing so.[33] Consequently, while the action does benefit us, its benefit is outweighed by the countervailing harms incurred from the desires that it

[32] Mark Schroeder uses the term 'proportionalism' to express the view that the strength of our reasons is always proportional to the strength of our desires (2007, 164–170). I am appropriating that term and applying it to the claim that the strength of our motivations is always proportional to the strength of our desires. The rejection of proportionalism plays an important role in my explanation of the harm of severe depression in §3.2 and §3.4. Another type of proportionalism is discussed in §4.2.

[33] Neil Sinhababu provides a supplementary explanation of the phenomenology of obligation. He points out that aversions may explain motivation in these cases rather than desires (Sinhababu 2017, 48). If this is correct, then the leaflet deliverer's aversion to falling short morally may be what motivates their self-sacrificial action. Avoiding the frustrations that aversions produce may not generate the same feelings of satisfaction that usually emerge from fulfilling a desire. This explains why self-sacrificial actions have a different phenomenological quality. While this approach can explain the phenomenology of some cases of self-sacrifice, I nevertheless think that we ought to supplement it by rejecting proportionalism (which Sinhababu does). If aversions were the sole motivator of self-sacrifice, then we ought to expect feelings of relief to accompany acts of self-sacrifice. This is not generally the case. Moreover, many, if not all, acts of self-sacrifice are supererogatory (Smilansky 2021, 2–3). Consequently, it would be surprising if such acts were always motivated by aversions to falling short morally.

frustrates. If we accept this view, then there is no need to appeal to the frustration of future desires or the deprivation of future desire fulfilments to explain every act of self-sacrifice. This approach marks an improvement on the existing explanations of how desire theories of well-being can capture the harm of self-sacrifice.

Rejecting proportionalism requires making a distinction between motivation and desire. This distinction has ample precedent in moral psychology. For instance, one view claims that motivation is a disposition towards an action. This may include a disposition towards doing and thinking about doing an action (Gregory 2021, 30). This means that under the right conditions we would think about and undertake that action (Firth 1952, 320). In contrast, desires are not always connected with actions. It is possible to desire something that no possible action could affect. For instance, I may desire that the English football team wins a major trophy again, while being in no position to affect that outcome. Moreover, while it is conceivable that all motivation is explained in terms of desire, it is at least conceptually possible that things other than desires could motivate us. This possibility is what undergirds the debate between Humeans and non-Humeans (§2.2.3). Consequently, distinguishing between desire and motivation in this way is uncontroversial. What remains to be established is whether there are independent reasons to think that desires sometimes motivate disproportionally to their strength. It is this question that I turn to in the next section.

2.4 Proportionalism about desire and motivation

I have argued that rejecting proportionalism can serve as the basis of a response to the problem of self-sacrifice. Our consequent understanding of self-sacrifice makes it recognisable as a harm by desire theories of well-being. Moreover, this explanation captures the phenomenology of some cases of self-sacrifice better than existing proposals within the literature. I turn now to consider the viability of rejecting proportionalism. I argue that there are compelling reasons for doing so. §2.4.1 examines how rejecting proportionalism can enrich our understanding of other psychological phenomena and increase the viability of other attractive philosophical theories. §2.4.2 provides an example of an alternative moral psychology to proportionalism. §2.4.3 placates the worry that rejecting proportionalism may entail controversies in our wider moral psychology.

2.4.1 Additional counterexamples to proportionalism

Alongside self-sacrifice, there are other psychological phenomena where the strength of motivation does not seem to be proportional to the strength of desire. Weakness of will is one such example (Gregory 2021, 34–35). Perhaps the most common variety of this occurs when we fail to defer gratification (Sinhababu 2017, 38). In such cases we sometimes appear to pursue weaker desires that motivate us with disproportional strength. After experiencing weakness of will, we often reflectively feel that we have acted against our presently strongest balance of desires. Moreover, in cases of 'clear-eyed' weakness of will, we may even feel that we are frustrating our presently strongest balance of desires as we act. However, a commitment to proportionalism forces us to accept that, despite appearances to the contrary, we are motivated by our strongest desires in these cases. If we reject proportionalism, then we can claim that some cases of weakness of will are caused by weaker desires that motivate us with disproportional force. This better captures the phenomenology of these cases.

Another case emerges when we forget our desires and they thereby fail to motivate us (Gregory 2021, 34). For instance, we may forget our desire to pay off a credit card before incurring fines. In this type of case, it appears that our desire has failed to motivate us because we did not keep it in mind when making decisions. Those who accept proportionalism must presumably claim that these desires have been lost and later reacquired when we remember them. However, this description is counterintuitive. The existence of non-occurrent beliefs is well-established within moral psychology. Given that desires share a similar structure to beliefs (Gregory 2012; Sumner 1996, 124), we should not be surprised that non-occurrent desires also exist. Indeed, if we want to capture the intuition that desires can persist through states of sleep and unconsciousness, then we need to assent to the existence of non-occurrent desires. We should be reluctant to accept a theory that does not recognise that desires continue to exist in those states. The persistence of desire over time is a big part of folk concepts of personal identity. It would have counterintuitive implications for these concepts if it turns out that desires are reconstituted every time we wake up. For these reasons, it is more intuitive to accept that non-occurrent desires do exist. If this is correct, then proportionalism is false because these desires do not motivate.

Additionally, some desires are unusually strongly felt. This is often the case with desires that are accompanied by strong emotions (Raibley 2010, 598–599). Anger often seems to amplify the motivational force of desires (Nussbaum 2016, 96). For instance, road rage can cause fleeting but strongly felt desires to shout at other motorists. In such cases, we are reticent to describe our desires as strong, as they quickly dissipate when our attention shifts. A more

intuitive description of the psychology of these cases claims that they involve relatively minor desires that motivate us with disproportional strength. The amplification of the motivational force of some desires may also be something that happens during episodes of mania. These tend to involve the restructuring of the relative importance of our desires (Porter 2023, 65–66). Nevertheless, when these episodes dissipate, desires often return to their previously perceived levels of importance. Rejecting proportionalism allows us to claim that an effect of mania is the amplification of the motivational force of some desires. It seems likely that other emotions and disorders similarly have amplificatory or attenuative effects on the motivational force of desires. Indeed, in the next chapter I argue that an effect of depression is the suppression of the motivational force of desires (§3.2).

The existence of self-sacrifice, weakness of will, forgotten desires, and unusually strongly felt desires undermine the attractiveness of proportionalism. Taken individually none of these are decisive arguments against this view. There may well be plausible characterisations of some of them that do not require its rejection. Nevertheless, these examples collectively illustrate the extent to which rejecting proportionalism can enrich our understanding of a range of psychological phenomena.

Aside from the psychological phenomena discussed, our position on proportionalism also has ramifications for our other philosophical views. For instance, some notions of moral responsibility are premised upon the idea that we have the capacity to choose not to act upon our strongest desires (Schoeman 1978). These theories draw their plausibility from the observation that it is counterintuitive to hold someone morally responsible for something that they could not have chosen to do otherwise. If we find this type of theory compelling, then a commitment to proportionalism entails that we are not morally responsible for our actions. This is because we are simply motivated by whichever desire happens to be strongest. However, if we reject proportionalism, then we can accept a theory of moral responsibility that claims that we are not powerless to resist our desires. This argument rests upon the idea that there are things we can do to control the motivational force of our desires. If this is right, then persons can be morally responsible for their actions. If we find a theory of this sort compelling, then this is an additional reason to reject proportionalism.

Relatedly, it may be that our concept of personhood relies upon the possibility of us having some agency over the shaping of our own desires. One way that such agency could exist is if we are able to have some control over the extent of the motivational force of our desires. For instance, Harry Frankfurt's influential concept of personhood seems to rest upon the idea that we are able to shape our own desires to at least some extent. For him, persons are those

beings that have second-order desires. These are the desires that we have about our own desires (Frankfurt 1971, 9). He acknowledges that, through deliberation, we are empowered to choose to act upon our desires in such a way that gives us some control over which in our wider polyphony of desires successfully motivates us (Frankfurt 1971, 11). For Frankfurt, freedom of the will requires that our second-order desires are for the action that our first-order desires successfully motivate us to perform (1971, 13). On his view, freedom of the will is a feature that only persons possess (Frankfurt 1971, 14).

Frankfurt does not explain how we can achieve agency over our desires. Nevertheless, one way of explaining how this can happen is to claim that there are things that we can do to amplify or attenuate the motivational force of our desires. Without a mechanism like this, it would appear that freedom of the will is something that we are powerless to pursue or protect. It would simply be good fortune if our second-order desires aligned with those first-order desires that successfully motivate us to action. Such a view is ill-befitting of the name of freedom of the will. However, if we postulate that we have some control over which desires motivate us, then this leaves some role for agency in our conception of freedom of the will. Consequently, it seems that Frankfurt's approach to personhood is also made more plausible by rejecting proportionalism. While these theories of agency and personhood are not the only ones available, nor are these specific formulations immune to critique, they are nevertheless plausible starting points for developing wider theories. Consequently, rejecting proportionalism has implications for our wider understanding about the type of creatures that we are.

2.4.2 An alternative moral psychology

One attraction of proportionalism is that it explains how strength of motivation is determined. On this view, strength of motivation is entirely proportional to strength of desire. While rejecting proportionalism allows us to provide more intuitive accounts of a range of psychological phenomena and related philosophical theories, it may also leave us bereft of a moral psychology that explains the relationship between desire and motivation. A theory with some counterintuitive implications may be more attractive than no theory at all. Consequently, it is incumbent upon those rejecting proportionalism to at least gesture towards an alternative.

Neil Sinhababu outlines one such alternative. He accepts that strength of motivation is roughly correlated with strength of desire (Sinhababu 2017, 3). However, following Hume,

he postulates that the phenomenon of mental vividness can amplify the motivational effects of desire (Sinhababu 2017, 36). It may be that the exercise of mental capacities, such as imagination, play a role in increasing mental vividness (Sinhababu 2017, 54). Intuitively, it seems that focussing our attention on certain desires can increase our motivation to fulfil those desires. Similarly, failing to pay attention to certain desires or aversions may also diminish their motivational force. This view allows for the motivational force of desires to fluctuate with attention. Sinhababu also thinks that the exercise of willpower can shift our attention towards and away from certain desires (2017, 135–145). This gives us some agency in controlling the motivational force of our desires. It also makes conceptual space for attractive theories of moral responsibility, personhood, and freedom of the will.

Sinhababu's view provides a viable alternative to proportionalism. It can intuitively account for cases of self-sacrifice, weakness of will, forgotten desires, and unusually strongly felt desires. On this view, these cases involve motivation being amplified or attenuated by the amount of attention paid to desires and aversions. This better captures the experiential quality of these psychological phenomena. Moreover, it is compatible with the Humean Theory of Motivation. On Sinhababu's view, motivation is contingent upon the existence of desire. However, it ceases to always be proportional to the strength of desire. Finally, by claiming that the strength of motivation roughly tracks the strength of desire, he captures some pro-proportionalism intuitions about its plausibility in everyday cases while accounting for the phenomenology of problem cases.

An advantage of this type of alternative to proportionalism is that it can preserve the Humean Theory of Motivation, while capturing some of the key intuitions of non-Humean views. For instance, moral beliefs are sometimes said to have motivational force independently of desires (§2.2.3). Rejecting proportionalism allows us to preserve the claim that only desires motivate, while acknowledging that moral beliefs can augment the extent of the motivational force of desires. The mechanism by which moral beliefs can affect motivational strength is through the exercise of attention. This increases the mental vividness of our desires to behave morally. This approach explains why moral beliefs sometimes fail to motivate, while acknowledging that such beliefs do often play an important role in motivation. According to this picture, when moral beliefs fail to motivate it is because we lack a desire with sufficient motivational force to act upon them. Conversely, when moral beliefs do appear to motivate us, they are instead amplifying the motivational force of the desires that we have to behave morally. They do so by drawing our attention to certain desires and away from others. Therefore, the rejection of proportionalism facilitates the adoption of a Humean position that

placates some non-Humean intuitions about motivation. It also explains how moral beliefs can amplify the motivational force of desires in some cases of self-sacrifice.

2.4.3 Implications for our wider understanding of desire

Some readers may be concerned that rejecting proportionalism requires taking controversial positions within philosophy of desire. If this is the case, then my solution to the problem of self-sacrifice will depend upon whichever theory of desire happens to be correct. Earlier, I dismissed the idea of rejecting the Humean Theory of Motivation to solve the problem of self-sacrifice partly because it would commit us to taking strong positions on controversies within moral psychology (§2.2.3).

It is true that ultimately my solution to this problem within philosophy of well-being depends upon findings within moral psychology. Nevertheless, the rejection of proportionalism is far less controversial than the rejection of the Humean Theory of Motivation. Rejecting proportionalism is compatible with a range of different theories of desire. If one takes desire to be characterised by a disposition to experience pleasure when imagining fulfilling desires (Strawson 2010), then there is no reason why we need to commit to proportionalism. If one takes desire to be primarily characterised by reward and learning responses (T. Schroeder 2004), then there is no reason why we need to commit to proportionalism. If one takes desire to be a subset of our normative beliefs (Gregory 2021), then there is no reason why we need to commit to proportionalism. Indeed, the only theories of desire that are in tension with the rejection of proportionalism are a subset of motivational views of desire.

According to motivational theories of desire, desires are defined by their effects on motivation. The most reductive of these theories claim that desires simply *are* motivation (Dancy 2000, 85). If this is correct, then proportionalism is trivially true. However, the reductive view is plainly incorrect. We often desire things that fail to motivate us because we lack a means-end belief to attain them. Moreover, our motivation is often dampened when we predict that our actions have only a limited chance of fulfilling our desires. In both of these cases, desires remain constant, while motivation fluctuates according to our beliefs. However, the reductive view counterintuitively entails that our desires are weakened or eliminated. Consequently, this view is implausible. Alternative motivational accounts of desire claim that desires are dispositions to be motivated towards outcomes. A disposition to be motivated does not entail that every instance of desire will motivate us; only that under

the right conditions desires will motivate. Therefore, this position is compatible with the rejection of proportionalism.

2.5 Chapter summary

Desire theories of well-being are sometimes said to be false because they fail to account for the harm of self-sacrifice. Chris Heathwood has shown that these theories can capture this harm. He appeals to the frustration of future desires and the deprivation of future desire fulfilments in order to do so. However, it is implausible to apply this explanation to every act of self-sacrifice. I have argued that if we reject proportionalism, then we can explain the harm of some cases of self-sacrifice as arising when we act upon our weaker desires and thereby frustrate our presently strongest balance of desires. This explanation better captures intuitions about the experiential quality of some cases of self-sacrifice. It therefore marks an improvement on Chris Heathwood's position. In the next chapter we will see how rejecting proportionalism also forms part of the picture of how desire theories of well-being can account for the harm of depression.

Chapter 3 The Problem of Depression

3.0 Introduction

This chapter considers a relatively new objection to desire theories of well-being. Ian Tully and Andrew Spaid have recently independently argued that these theories entail that severe depression does not diminish well-being. If this is correct, then desire theories of well-being are undermined by the counterintuitive nature of their implications. I argue that proponents of these theories have two supplementary ways of capturing the harm of most cases of severe depression. Firstly, they can point out that severe depression reduces motivation while leaving some desires intact. This leads to the frustration of those desires that we are left unmotivated to fulfil. Secondly, they can observe that severe depression prevents pleasure from emerging when pursuing our desires. This inevitably frustrates desires that specify pleasure in their content. These explanations can account for the harm of the majority of cases of severe depression. Nevertheless, depression is a heterogeneous phenomenon. Consequently, residual cases require additional explanations. I argue that sometimes the harm of residual cases of severe depression can be explained by self-destructive desires to be badly off. These are often born of the emotional turmoil and negative self-evaluative beliefs that some sufferers of depression experience. At other times, severe depression leaves us worse off than we would otherwise have been by preventing the formation and subsequent fulfilment of future desires. These four complimentary explanations of the harm of severe depression solve this problem for desire theories of well-being and enrich our understanding of how depression works.

This chapter has the following structure: §3.1 outlines the problem of depression for the desire theories of well-being. §3.2 argues that an effect of severe depression is that it decreases motivation while leaving some desires intact. §3.3 argues that another effect of severe depression is the inevitable frustration of desires that specify pleasure in their content. §3.4 examines residual cases of severe depression that are less well explained by these approaches. §3.5 considers the relationship between severe depression and prudential reasons. §3.6 concludes with a summary.

3.1 The problem of depression

The problem of depression for desire theories of well-being was introduced in a 2017 paper by Ian Tully. Another iteration of this problem was put forward in Andrew Spaid's 2020 PhD thesis. This chapter considers versions of the argument put forward by both authors that coalesce around the claim that desire theories of well-being are unable to account for the harm of severe depression. Tully argues that desire theories of well-being are unable to capture the intuition that severe depression is a state of harm.[34] To make his case, he appeals to a distinction between two species of depression: motivational and consummatory anhedonia (Tully 2017, 4). The former occurs when we lose interest in pursuing many of the things that we previously enjoyed. Whereas the latter occurs when we consistently fail to experience pleasure from satisfying our desires. He argues that the clearest cases of severe depression emerge when someone experiences both concurrently. According to Tully, this symptomology indicates that our desires have been weakened or eliminated. On this view, motivational anhedonia is taken as evidence that we lack desire; whereas consummatory anhedonia is taken as evidence that we did not desire the outcome in the first place.[35]

For the purposes of this chapter, we can loosely define severe depression as constituted by experiencing both types of anhedonia concurrently and to a significant degree. However, it is not necessary to commit to any particular nosology of depression in order for the arguments of this chapter to pose a challenge to desire theories of well-being. It would be a problem for these theories if they were unable to capture the harm of experiencing significant amounts of consummatory and motivational anhedonia concurrently, regardless of whether we assent to describing every instance of doing so as a case of severe depression. This caveat is important because there are significant debates as to what qualifies a condition as depression. For instance, on contextual views, the origins of depressive symptoms play an important role in determining whether a condition qualifies as depression (Tully 2019). According to these views, depressive symptoms that are appropriate and proportionate responses to external factors do not qualify a condition as depression. This type of approach is often motivated by a desire to avoid the pathologisation of ordinary sadness in response to things such as bereavement, unemployment, or immiserating social arrangements (Hari

[34] Tully refers to states of ill-being rather than states of harm. However, for simplicity, I will refer to states of harm instead. In §1.1 I argued that there are good reasons to conceptualise ill-being as negative well-being, rather than as a distinct value in itself.

[35] The view that depression weakens or eliminates desires has substantial philosophical precedent (Smith 1994, 135; Stocker 1979, 744).

2018). Other views claim that depressive symptoms, when experienced significantly, are enough to qualify a condition as depression irrespective of the cause of those symptoms (Kendler et al. 2010). There are a number of reasons why one might hold this view. For instance, sometimes this view is held because of the difficulty in identifying when sadness is appropriate and proportionate. I set aside this debate here as it is of no consequence to the argument of this chapter.

Viable theories of well-being must have intuitive implications about paradigm cases of benefit and harm. Undoubtedly, severe depression is a paradigm case of harm. However, Tully argues that desire theories of well-being fail to entail this claim. This is because, on these views, for something to diminish well-being it must frustrate desires. If severe depression weakens or eliminates desires, rather than frustrates them, then it is not a state of harm. I will refer to this argument as 'the problem of depression'.

The problem of depression can be summarised as follows:

P1: Severe depression is a state of harm that viable theories of well-being must capture.

P2: Desire theories of well-being entail that severe depression is not a state of harm.

C: Therefore, desire theories of well-being are not viable theories of well-being.

This argument is valid. Moreover, premise one is overwhelmingly intuitive. However, premise two is contestable. If it can be shown to be incorrect, then desire theories of well-being can avoid this problem.

3.2 Understanding motivational anhedonia

One way of responding to the problem of depression involves challenging the characterisation of severe depression that it is premised upon. Tully takes the coexistence of consummatory and motivational anhedonia to indicate lack of desire. There is something intuitive about this explanation. If both pleasure and motivation are absent, and there is no countervailing evidence, then it seems reasonable to assume that desire is absent. If this is right, then desire theories of well-being struggle to account for the harm of severe depression.

However, there is good reason to think that this characterisation of severe depression is incomplete. This is because it fails to capture some aspects of its experiential quality. Testimonial evidence illuminates this. Severe depression is often described as suppressing motivation, while leaving underlying desires intact. Often severely depressed people report

having desires but being unable to motivate themselves to fulfil them. In the words of Steven Swartzer, 'This is part of what is so frustrating about such experiences. That one is unable to engage in activities that one cares strongly about is part of why such situations are so heartbreaking' (2015, 9, underlining added).[36]

To illustrate this point, Swartzer appeals to the testimony of a man who is unable to motivate himself to attend his son's wedding:

> 'I knew that my son's wedding would be emotional … and that anything emotional, good or bad, sets me off. I wanted to be prepared. I'd always hated the idea of electroshock therapy, but I went and had it anyway. But it didn't do any good. By the time the wedding came, I couldn't even get out of bed. It broke my heart, but there was no way that I could get there' (Swartzer 2015, 8).

In this case, the man's desire to attend the wedding appears to persist, even though it fails to motivate him to action. Given his testimony, it seems perverse to insist that the man did not *really* desire to attend his son's wedding. Consequently, there are good reasons to think that severe depression has harmed him by dampening the motivational force of his desires and thereby frustrating them. On this view, motivational anhedonia is explained as an aspect of severe depression that suppresses motivation directly.

Nevertheless, one might concede a strong desire on the man's behalf but postulate stronger countervailing aversions. If aversions have a different phenomenological character to desires (Sinhababu 2017, 48), then the emotional turmoil reported in Swartzer's example may be explained by them. It may be that the man in Swartzer's example fails to disclose these aversions because he does not identify with them. Nevertheless, the failure to identify with these aversions does not mean that they are not relevant to his well-being.[37] Consequently, perhaps his strongest desires are fulfilled by continuing to languish in bed, rather than by

[36] Other writers characterise depression similarly (Arpaly & Schroeder 2014, 126; T. Schroeder 2004, 31–32).

[37] It has been pointed out that some mental states seem to not truly be a part of us in the same way that others are (Penelhum 1979). Some versions of the desire theory of well-being claim that desires which fail to resonate with our deepest concerns are not constitutive of our well-being (Noggle 1999, 314–316). If this type of theory is correct, then it may be that the aversions postulated in response to Swartzer's example do not improve well-being when fulfilled. Consequently, even if the man's inertia is explained by his aversions, fulfilling these aversions does not benefit him. Conversely, he is harmed by the frustration of his desires to attend the wedding because those desires do resonate with his identity. My argument does not rely upon this type of restriction. In §4.4 I consider how desire theories of well-being can account for desires that seem to be alienated from our identity without restricting which desires affect well-being.

attending his son's wedding. If this is right, then severe depression does not harm him by obstructing the fulfilment of his strongest desires. Therefore, desire theories of well-being seem unable to recognise severe depression as an aggregate harm in this case.

However, it strikes me that an appeal to aversions less convincingly captures the phenomenology of Swartzer's example than the suppression of the motivational force of desires. Postulating unreported aversions in Swartzer's example is a less parsimonious explanation of what is happening than simply accepting the man's testimony that his desires are failing to motivate him to action. Moreover, we have seen that there are other psychological phenomena that involve desires motivating disproportionally to their strength (§2.4.1). Given this precedent, it should not be too surprising that depression is part of the pantheon of psychological phenomena where desires motivate disproportionally to their strength. It is possible to construct additional examples that more clearly illustrate this effect. This is more apparent in cases of severe depression that do not involve the emotional turmoil present in Swartzer's example. For instance, sufferers of severe depression sometimes struggle to motivate themselves to get out of bed despite having many desires whose fulfilment is dependent upon doing so. In such cases, it is more intuitive to claim that desires are failing to motivate, rather than being outweighed by countervailing aversions.

These types of cases are counterexamples to Tully's characterisation of severe depression. They illustrate how his view fails to capture an aspect of severe depression's experiential quality. The most intuitive way of rectifying this position is to claim that one effect of severe depression is that it leaves some desires intact, while dampening their motivational force. On this view, severe depression makes us less able to fulfil our desires. Consequently, it is a state laden with desire frustrations. If this is right, then desire theories of well-being can explain the harm of severe depression as arising from the desires that it frustrates. This characterisation of severe depression is more economical than Tully's view. He infers lack of desire from the existence of motivational and consummatory anhedonia. Conversely, I take motivational anhedonia to indicate a lack of motivation. This argument is premised upon the view that desires sometimes fail to motivate proportionally to their strength. Therefore, it requires the rejection of proportionalism about desire and motivation. In §2.4.1 I argued that there are strong independent reasons to reject that view.

3.3 Understanding consummatory anhedonia

Tully anticipates the argument that severe depression may dampen motivation while leaving desires intact. One way of making this explanation congruent with the phenomenology of severe depression is to claim that severe depression masks desires (Tully 2017, 10–12). Masked desires are undetectable through introspection and do not motivate. One way in which this may happen is if severe depression prevents the dispositions that are constitutive of desire from manifesting (Tully 2017, 11). Consequently, perhaps severe depression does not weaken or eliminate desires in the way that the problem of depression claims. If it instead masks desires, then desire theories of well-being can recognise it as a state of harm. This is because it inevitably frustrates the desires that it masks. In defence of the masked desires hypothesis, we can point to the existence of forgotten desires as a precedent for non-occurrent desires (§2.4.1). We can also observe that when we are asleep or unconscious it seems that we retain desires that do not manifest. Perhaps severe depression affects desires similarly.

However, Tully finds this type of argument unpersuasive for two reasons. Firstly, he points out that if desires are masked by depression, then desire theories of well-being implausibly entail that their fulfilment improves well-being (Tully 2017, 12). To illustrate this point, he constructs an example of having a favourite meal prepared for dinner while severely depressed (Tully 2017, 11–12). He takes the absence of feelings of satisfaction when eating the meal to be evidence that well-being is not improved. Consequently, even if severe depression does mask desires, this will not rescue desire theories of well-being from their own counterintuitive implications. Secondly, Tully points out that when depression dissipates, not all desires return. Indeed, some etiological theories of depression claim that part of its evolutionary function is to divest us of those desires that are not in our best interests to retain (Tully 2017, 11). Tully takes these arguments to be evidence of the view that severe depression weakens or eliminates desires, rather than suppresses their motivational force.

The argument that I made in §3.2 does not rely upon the existence of masked desires. It is perfectly plausible to claim that an effect of severe depression is that it leaves us painfully aware of those desires that we are left unmotivated to fulfil. The testimony appealed to in Swartzer's example illustrates this. Since I am not seeking to defend the idea that severe depression masks desire in this way, Tully's argument does not refute my position. Nevertheless, I will address Tully's argument on his own terms.

There are at least three responses to Tully's first objection. Firstly, we can point out that desire theories of well-being do not require feelings of satisfaction to be present for well-

being to be improved (§1.2). Desires can be fulfilled without generating these feelings. For instance, this happens when we do not find out about our desire's fulfilment. Nevertheless, while feelings of satisfaction are not required for desire fulfilments to improve well-being, the lack of these feelings is often a good indicator that we have not fulfilled our desire. Consequently, without supplementation, this response to Tully's example is unconvincing.

A second response is to interpret examples such as Tully's as not describing genuine desire fulfilment, but rather mere simulacra. Desires can be more complex than is sometimes assumed. For instance, sometimes desires are conjunctive.[38] A conjunctive desire is one that can be accurately represented as having multiple propositions in its content which are conjoined by an "and" logical operator. In order for a conjunctive desire to be fulfilled, all of its conjuncts must be true. It is possible to redescribe Tully's example as involving a conjunctive desire. For instance, it may be that the desire is to *enjoy* the favourite meal. Desires for enjoyment have two distinct propositions in their content. They specify an object, and they specify that the subject with the desire takes pleasure in the existence of that object. If the desire that Tully refers to is of this type, then it specifies having his favoured meal prepared for him *and* taking pleasure in eating it. If severe depression prevents pleasure from emerging, then this inevitably frustrates desires of this sort. This is because it makes one of the conjuncts impossible to fulfil. Consummatory anhedonia is built into Tully's definition of severe depression (Tully 2017, 4). Consequently, he acknowledges that one effect of depression is the prevention of pleasure from emerging. However, he fails to appreciate the extent to which some desires have both pleasure and other objects in their content. If this is right, then desire theories of well-being can explain why we are not always benefitted by the apparent fulfilment of masked desires. This is because some desires cannot be fulfilled without pleasure being present, and consummatory anhedonia prevents pleasure from emerging.[39]

On this view, pleasure is built into the fulfilment conditions of some desires. It does not take great powers of introspection to reveal that many of our desires have this structure. This often seems to be true of our everyday desires. For instance, many of our desires for food,

[38] Evan G. Williams briefly notes that some desires are conjunctive before discussing a different issue (2017, 213). I discuss conjunctive desires further in §5.3 and §5.4.

[39] This analysis of consummatory anhedonia also sheds light on the relationship between both species of anhedonia. If we anticipate that consummatory anhedonia makes many of our desires unfulfillable, then we are unlikely to be motivated to pursue those desires in the future. This creates a feedback loop whereby experiences of consummatory anhedonia fuel increased motivational anhedonia.

sex, and leisure are structured in this way. These desires specify that we must experience pleasure from the activity in order for our desire to be fulfilled. It is revealing that Tully's example of a masked desire that does not improve well-being when fulfilled is of this sort. However, not every desire has this structure. For example, when we write our last will and testament, we are expressing our desires for events that we want to happen after our death. These desires cannot specify pleasure in their content unless they are very confused in their structure. After all, we will not be alive to experience any pleasure in response to observing the fulfilment of our posthumous desires. Moreover, many of our long-standing desires are less concerned with pleasure than their everyday counterparts. For instance, consider your desire to maintain a close friendship, to finish writing your masterpiece, or to live in accordance with your deeply held values. It seems less clear that we would not benefit from the fulfilment of these desires if consummatory anhedonia were to block pleasure from emerging. Consequently, we ought not to overgeneralise the observation that some desires specify pleasure in their content.

Nevertheless, this argument is ill-suited to account for all cases of severe depression. If we cannot introspectively detect the existence of a desire, then often the most intuitive explanation is to accept that no desire is present. Consequently, a third response to Tully's argument is to accept that severe depression has eliminated the desire in his example. This means acknowledging that severe depression sometimes weakens or eliminates desires. This also explains why not all desires return when depression lifts. Although, we should note that desires do change over time anyway. The fact that people emerge from severe depression with different desires does not undermine the argument that severe depression masks desires. It may simply have masked desires that gradually changed throughout the course of depression in the same way that desires gradually change in those without depression. Accepting that desires can be eliminated or weakened by severe depression does not mean that desire theories of well-being cannot recognise it as a state of harm. On this view, the removal of some desires is a benign effect of severe depression, while its harm emerges from the desires inevitably frustrated by their inability to motivate and by its prevention of pleasure from emerging.

3.4 Residual problem cases

I have argued that proponents of desire theories of well-being can recognise severe depression as a state of harm. I have discussed two ways in which they can do so. Firstly,

they can point out that severe depression dampens motivation while leaving some desires intact. This leads to the frustration of those desires that we are left unmotivated to fulfil. Secondly, they can claim that severe depression prevents pleasure from emerging when pursuing desires. This inevitably frustrates desires that specify pleasure in their content. It strikes me that these explanations are a more intuitive way of characterising motivational and consummatory anhedonia than Tully's inference that these effects indicate a lack of desire. It also seems likely to me that these explanations can account for the harm of most cases of severe depression. Nevertheless, depression is a heterogenous phenomenon. Consequently, there remain problem cases that are less well explained by this analysis. I examine two types of problem cases in this section. The first are cases of severely depressed people who seem to have very few desires that are almost all fulfilled. The second are cases of severely depressed people who seem to be completely devoid of desires. Desire theories of well-being should have something to say about these residual cases. I consider them sequentially.

Andrew Spaid provides a fictionalised example of a woman called Jane:

> 'Jane is diagnosed with clinical depression, and understands that she is depressed. She also understands that an effective treatment for her depression is available. In other words, she understands that with treatment she would come to have the desires and the joys most non-depressed people have—in short, a normal life. Nevertheless, Jane refuses treatment for her current episode of depression, claiming that she does not care about the treatment outcome—she sees no point in regaining the desire to live because she believes nothing is worth doing' (Spaid 2020, 28–29).

Jane appears to lack a great many desires. Spaid speculates that the few desires that she retains are fulfilled. Consequently, desire theories of well-being seem to entail that Jane is not badly off. After all, she does not appear to want things that she does not have. Therefore, these theories seem forced to conclude that her well-being is at least mildly positive. Yet, this is an unintuitive implication. Her listlessness is unenviable. Desire theories of well-being should be able to explain why we would be reticent to swap places with her. Spaid supplements this fictionalisation with a wealth of supportive testimony from sufferers of depression that similarly illustrate his point (2020, 38–39).

One way of responding to cases like this is to claim that people like Jane have self-destructive desires to be badly off. On this view, the fact that Jane refuses treatment is not taken as evidence of an absence of desire. Rather, it is taken as evidence of a self-destructive desire to remain badly off. This explains why Jane actively refuses, rather than simply expresses

indifference to, treatment. If this is right, then desire theories of well-being can recognise Jane's severe depression as a state of harm. This is because Jane has a self-destructive desire to remain badly off that motivates her with disproportional strength to obstruct the fulfilment of her other stronger desires.[40] Accordingly, she has many frustrated desires.

I do not think that cases like Jane's are particularly common. Most severely depressed people do not have desires to be badly off. Indeed, the majority of people with this condition will go to great lengths to try to alleviate their depression. The man in Swartzer's example is not unusual in this respect. Moreover, the example of having a general desire to be badly off is somewhat confected. It seems that, rather than having general desires to be badly off, it is far more common for severely depressed people to have some specific self-destructive desires that obstruct the fulfilment of their other stronger desires. It may be that these self-destructive desires to be badly off are born of the negative self-evaluative beliefs that some sufferers of severe depression experience (Spaid 2020, 38; Goldman 2009, 105). Depression can be merciless at drawing our attention to these desires through the emotional turmoil that it cultivates. As we have seen, it seems likely that attention has amplificatory effects on motivation (§2.4.2). Consequently, it may be that these self-destructive desires to be badly off are motivating with disproportional strength because of the attention that severe depression draws to them. This leads to the frustration of other desires that have cumulatively greater strength. If this is right, then Spaid's characterisation of Jane is incorrect. Jane does not have very few desires which are all fulfilled. Rather, she has many desires that her self-destructive desires to be badly off successfully motivate her to frustrate. The acquisition of such desires seems to be an effect of severe depression that some people experience.

Nevertheless, Spaid is probably right that there are some severely depressed people who have very few desires, the majority of which are fulfilled. If this is right, then none of the approaches that I have put forward succeed in capturing the harm of these residual cases.

[40] Readers may be concerned that if Jane fulfils her self-destructive desire to be badly off, then she is benefitted by this desire's fulfilment. Chris Heathwood responds to this concern by accepting that the fulfilment of desires to be badly off does provide some benefit. However, this benefit is outweighed by the harm of frustrating other desires. Consequently, fulfilling a desire to be badly off is an aggregate harm (Heathwood 2005, 501–502). Nevertheless, a paradox for desire theories of well-being can emerge when the fulfilment of a desire to be badly off generates enough well-being to make a person's life overall good for them (Bradley 2007). In this case, if Jane's desire to be badly off is fulfilled, then she would have positive well-being, which would then mean that her desire is frustrated, which would mean that she has negative well-being, which would mean that her desire is fulfilled, and so on. I set aside this problem here. Other writers have proposed potential solutions to this problem (Dorsey 2012b, 422–425; Skow 2009).

Spaid cites Tolstoy's autobiographical account of his depression as illustrative of this type of case (Spaid 2020, 39):

> 'My life came to a standstill. I could breathe, eat, drink, and sleep, and I could not help doing these things; but there was no life, for there were no wishes the fulfillment of which I could consider reasonable. If I desired anything, I knew in advance that whether I satisfied my desire or not, nothing would come of it. Had a fairy come and offered to fulfill my desires I should not have known what to ask…I could not even wish to know the truth, for I guessed of what it consisted. The truth was that life is meaningless' (Tolstoy 1954, 17).

At least for the sake of argument, we can accept Spaid's interpretation of this passage as illustrating that the primary symptom of Tolstoy's condition is the elimination of his desires. Crucially, we should interpret Tolstoy as not experiencing any emotional distress or anguish. This is because it seems likely that human psychology is structured in such a way as to have standing desires to be free from mental turmoil (§1.3). If Tolstoy were experiencing aversive emotions, then this would serve as evidence that these standing desires are frustrated. Consequently, our understanding of Tolstoy's severe depression should be that he simply does not have many desires, and those that he has are largely fulfilled. Desire theories of well-being should have something to say about such cases.

Proponents of these theories can appeal to a deprivation account to explain why people like Tolstoy are made worse off by severe depression. Some views about the harm of death take this approach (Feldman 1991; Nagel 1970b). On these views, death is bad because it deprives us of future goods. Consequently, it is not an intrinsic harm but rather an instrumental or extrinsic harm (Bradley 2008, 300). We can apply a similar argument to explain the effects of Tolstoy's condition on his well-being. On this view, Tolstoy is instrumentally harmed by his depression because it makes him less able to generate and fulfil new desires. It therefore leaves him worse off than he would otherwise have been.

The deprivation of future goods is surely an effect of severe depression. However, deprivation is not an intrinsic harm. According to desire theories of well-being, to be made badly off requires having frustrated desires (§1.2). Consequently, it is difficult to see how Tolstoy's severe depression has made him badly off, rather than simply worse off than he otherwise would have been. While deprivation accounts can explain why people with severe depression lack positive well-being, these accounts fail to explain how this condition leaves them with negative amounts of well-being. Therefore, this approach does not adequately account for the harm of most cases of severe depression.

Nevertheless, it goes some way to explaining residual cases. When we think of severe depression as a paradigm case of harm, we tend to imagine someone unmotivated to fulfil their desires, someone who fails to experience pleasure when pursuing their desires, or someone who is plagued by self-destructive desires that keep them badly off. Proponents of desire theories of well-being can explain why people who experience any of these effects are harmed by severe depression. Conversely, if we imagine someone truly devoid of almost all desires, and without other depressive symptoms, then it strikes me as less clear that this person is in a state of harm. Indeed, it is unclear to me whether such a person can accurately be described as having severe depression. After all, someone in this condition appears to be in a better position than someone with many desires who fails to experience motivation or pleasure, or someone beset with self-destructive desires that successfully motivate them to remain badly off. Consequently, our reticence to swap places with people in Tolstoy's position may stem from the fact that doing so would leave us significantly worse off than we are currently, rather than badly off. Even if our intuitions persist that Tolstoy is in a state of harm, rather than one of deprivation, then I think that these intuitions should not be considered decisive. Our criterion for adequate theories of well-being is that they give intuitive results in paradigm cases of benefit and harm (Fletcher 2016, 10). Atypical cases such as Tolstoy's are by no means paradigm cases of harm. Consequently, if this is a bullet to bite, then it is a far smaller one than Spaid's argument suggests.

The suppression of the motivational force of desires explains the harm of many cases of severe depression. The harm of others is explained by this condition's prevention of pleasure from emerging. The harm of still others is explained by self-destructive desires to be badly off that sometimes arise from the emotional turmoil and negative self-evaluative beliefs that some sufferers of severe depression experience. Moreover, an appeal to a deprivation account can explain residual cases as making us worse off than we otherwise would have been. Often these aspects of depression's harm exist to various extents in combination with each other. However, there may remain some cases that are ill-described by this analysis. Sufferers of complete conative collapse seem to fall into this category (Tully 2017, 6). Tully cites Viktor Frankl's description of some prisoners in Nazi concentration camps as illustrative of this state:

> 'The day would come when [they] would simply lie on their bunks in the barracks, would refuse to rise for roll call or for assignment to a work squad, would not bother about mess call, and ceased going to the washroom. Once they had reached this state, neither reproaches nor threats could rouse them out of their apathy. Nothing frightened

> them any longer; punishments they accepted dully and indifferently, without seeming to feel them' (Frankl 1986, 117).

Tully takes these cases to show that complete desirelessness is not only conceivable, but also has real world precedents. Crucially, people in this condition remain conscious and aware of their surroundings (Tully 2017, 6). An appeal to a deprivation account fails to sufficiently capture our intuitions about the well-being of people in this state. There are two reasons for this. Firstly, our theory of well-being should capture the intuition that people in this condition are badly off. An appeal to a deprivation account only succeeds in describing them as worse off than they would have otherwise been. Secondly, our theory of well-being should be able to recognise that people in this condition are subjects of well-being. However, desire theories of well-being typically claim that to be a subject of well-being one must have the capacity to have desires (§1.2). Consequently, if complete conative collapse eliminates this capacity, then desire theories of well-being implausibly entail that people in this condition are no longer subjects of well-being (Tully 2017, 7). Therefore, there is nothing that can be done to benefit or harm them. This is clearly an unacceptable implication.

One potential response to these problems is to claim that different standards of well-being are applicable to different types of being. Perhaps desire theories of well-being should only be used to explain well-being in subjects with the capacity to have desires, and another theory should be used to explain well-being in subjects who lack this capacity. However, this approach inherits the unenviable task of explaining which standards apply to which subjects (§1.4). Moreover, it forfeits one of the attractions of desire theories of well-being. This is their ability to explain which beings qualify as subjects of well-being (§1.2). Furthermore, by claiming that different standards apply to different subjects, this approach entails that the standards that determine a subject's well-being change when they acquire or lose the capacity to have desires. This means that their well-being could radically change for better or worse based on whether they have acquired or lost this capacity (Lin 2017a, 359–360). For these reasons, we should reject this modification to standard versions of the desire theory of well-being and look elsewhere to explain how complete conative collapse affects well-being.

A more promising approach is to reject Tully's characterisation of complete conative collapse as a state of complete desirelessness. We should accept that the primary effect of some non-standard types of severe depression is the elimination of a great many desires. Nevertheless, I remain unconvinced that severe depression could ever remove all of a subject's desires while leaving them sentient. It strikes me that part of having sentience is to

have at least some basic standing desires for certain positive mental states and at least some basic standing aversions to be free of other negative mental states (§1.3). If this is right, then sufferers of complete conative collapse have at least some basic desires that remain frustrated. The horrors that some victims of torture have endured may stifle the motivational force of these desires. This strikes me as a more plausible characterisation of complete conative collapse than complete desirelessness. After all, there are not any other obvious precedents of beings that lack all desires but nevertheless retain sentience. It would be strange if sentient desirelessness was only present in sufferers of complete conative collapse and nowhere else. Therefore, this is almost certainly not what is happening in these cases.

3.5 Prudential reasons and severe depression

My analysis of Tolstoy's severe depression raises wider questions about our understanding of prudential reasons. Given that Tolstoy's severe depression makes him worse off, it seems natural to think that he has a prudential reason to try and recover from his condition.[41] One way of capturing this intuition is to claim that people always have a prudential reason to try and prevent themselves from being made worse off. Accordingly, this line of reasoning entails that people who are made worse off through lack of desire have a prudential reason to acquire more desires if they are able to fulfil them. If this is right, then we have a prudential reason to try and recover from cases of severe depression that inhibit our ability to acquire new desires.

Unfortunately, this argument applies to non-depressed people as well. After all, according to desire theories of well-being, we are all made worse off by lacking more desires that would otherwise be fulfilled.[42] Dale Dorsey highlights the counterintuitive nature of this implication with an illustrative example:

> 'Faith is a highly regarded Air Force pilot who has long desired to become an astronaut. She has the physical skill, the appropriate training, and has been looked on

[41] This section is concerned with non-instrumental prudential reasons. I omit the non-instrumental qualifier to facilitate ease of readability.

[42] This is an implication of Heathwood's Desire Satisfactionism (2011, 24). In §2.2.4 I argued that desire theories of well-being must accept this view in order to account for the intuitions that we can be made worse off by fulfilling our presently strongest balance of desires. Alexander Dietz also argues that attempts to restrict which desires affect well-being to only our presently existing desires are implausible (2023).

> as a potential candidate. At time t, she has the choice to undergo the last remaining set of tests to become an astronaut or take a very powerful psychotropic pill that would have the result of radically, and permanently, changing her desires. Instead of preferring to be an astronaut, she could instead prefer to be a highly regarded, but Earth-bound, Air Force pilot' (Dorsey 2019, 161).

Clearly our theories of well-being and prudential reasons ought not to entail that Faith should be indifferent between these two outcomes. It is counterintuitive to accept that we have a prudential reason to acquire new desires simply to fulfil them, or to engineer our desires in such a way so that they better conform to the world as it is (Dorsey 2019; Barry 1989).[43] It is even more counterintuitive for our theory of prudential rationality to be indifferent between engineering our desires to fit the world and changing the world to fit our desires (Dorsey 2019, 162). Nevertheless, this seems to be an implication of accepting that people like Tolstoy have a prudential reason to try to recover from their severe depression. Consequently, we are left in a bind. Either we accept that people have prudential reasons to modify their desires in such a way as to maximise desire fulfilment, or we concede that people like Tolstoy have no prudential reason to try to cure their severe depression.

I think that we ought to opt for the second option. To my mind, it is intolerably counterintuitive for a theory of well-being to entail that we have a prudential reason to acquire new desires simply in order to increase our well-being. After all, our desires are constituent parts of our identity. It would be a significant affront to our capacity for self-authorship were we to be at the mercy of prudential reasons to maximise our well-being in this way. Moreover, we are almost always far more inclined to change the world to fit our desires, than we are to change our desires to fit the world. We cannot capture the intuition that we do not have a prudential reason to engineer our desires in this way by claiming that only presently existing desires matter to well-being (Dietz 2023; Heathwood 2011; §2.2.4). Consequently, we should claim that, while changing our desires to better fit the world does improve our well-being, we have no prudential reason to engineer our desires in this way

[43] It may be thought that therapy sometimes encourages us to acquire new desires in order to fulfil them. However, this strikes me as a mischaracterisation. Therapy tends to be concerned with encouraging us to divest ourselves of those desires which keep us badly off, with helping us better understand the content of our own desires, and with enabling us to identify those obstacles that prevent us from fulfilling our desires. Consequently, therapy is not a counterexample to the claims made in this section.

(Dorsey 2019).[44] On this view, we do not have a prudential reason to avoid making ourselves worse off by acquiring new desires that would be fulfilled. If this is right, then Tolstoy lacks a prudential reason to try to cure his severe depression. Admittedly, this implication is somewhat unsettling. Nevertheless, it strikes me as being less so than its alternative. Moreover, it is made more palatable when we keep in mind that we do not need to appeal to a deprivation account to explain why most cases of severe depression diminish well-being.[45]

This discussion may alarm some readers who are otherwise sympathetic to desire theories of well-being. After all, the idea that we always have a prudential reason to try to prevent ourselves from being made worse off seems intuitive. Many other theories of well-being can capture this intuition. Yet, desire theories of well-being cannot do so without leading to counterintuitive implications. For some readers this may serve as a reason to reject these theories. While I accept that this is somewhat unintuitive, I do not think that it fatally undermines the appeal of these theories. After all, proponents of these theories can claim that we have a prudential reason to avoid making ourselves worse off in many cases. They can also claim that we have prudential reasons to prevent ourselves from being made badly off. They need only claim that there is no such prudential reason to prevent ourselves from being made worse off by failing to engineer our desires to better fit the world. Moreover, other theories of well-being that claim that subjective attitudes such as desires, values, or goals are components of well-being inherit the same problem. These theories will need to make similar claims about prudential rationality if they are to avoid the implication that we have a prudential reason to engineer our subjective attitudes to maximise well-being. Consequently, the problem is not unique to desire theories of well-being.

There is a parallel between this discussion of prudential reasons and discussions about the ethics of creating new people. Many people have the intuition that we have no obligation to create new people with high well-being. Yet, these same people often have the seemingly contrasting intuition that a universe with more well-being is a better universe. At first glance, these two intuitions seem incompatible. Nevertheless, one way of reconciling them is to claim that, while creating more people with high well-being would improve the universe, we have no ethical reason to do so. On this view, ethical reasons grounded in beneficence do

[44] If we lack prudential reasons to always maximise our own well-being, then it would be profoundly strange were we to have ethical reasons to always maximise the well-being of others. Consequently, if the reader is sympathetic to this argument, then it ought to provoke some additional scepticism about the viability of act utilitarianism.

[45] It is possible that there are other reasons to try to recover from these cases of severe depression. For instance, perhaps the effects of severe depression are themselves an affront to self-authorship.

not mandate utility maximisation in all cases. The way in which utility is maximised is also relevant to the shaping of these reasons. Brian Barry puts this point like this:

> 'The proposition that it is a good thing for people to be happy does not entail that there is an obligation to bring into existence people who will be happy. Similarly, the proposition that it is a good thing for wants to be satisfied does not entail that there is an obligation to bring into existence easily satisfied wants' (Barry 1989, 281).

Here Barry is primarily discussing obligation and ethics, rather than prudential reasons and well-being. Nevertheless, his remarks make it clear that the discussions are parallel. If Barry is right, then ethical reasons do not always mandate utility maximisation. Similarly, if the argument that I have made is right, then prudential reasons are not solely concerned with well-being maximisation. The way in which we improve our well-being is also non-instrumentally relevant to the shaping of these reasons.

An alternative response to this problem claims that, while we always have a prudential reason to avoid being made worse off, this is often outweighed by reasons based other values, such as the value of self-authorship (Dietz 2023, 997; Barry 1989, 280–281). This explains why we generally do not try to improve our well-being by engineering our desires to better match the world as it is. Moreover, it does so without requiring us to revise our intuitions about prudential reasons. Nevertheless, I find this response a less convincing alternative. While an appeal to the value of self-authorship can explain why we do not have an all-things-considered reason to drastically change our desires in many cases, changing some of our desires can be done without radically changing our identity (Dorsey 2019, 163). For instance, for most of us, changing our desires for certain peripheral hobbies or leisure activities to ones that are easier to fulfil (perhaps by being cheaper or less time intensive) would not change our identity greatly. It is easy to imagine in such cases that the small affront to self-authorship of making this change could be outweighed by larger gains in well-being. Consequently, we should expect ourselves to sometimes have all-things-considered reasons to change our desires to better fit the world. Nevertheless, it still seems counterintuitive to accept that we have a prudential reason to make such changes. Most of the time we are completely indifferent between having new fulfilled desires or not having those desires. This approach fails to explain this preference.

A slightly different variation of this argument claims that we have countervailing prudential reasons to keep our desires relatively stable, rather than countervailing reasons based on other values (Yu 2022). For instance, Xiang Yu argues that it seems plausible that human psychology is structured in such a way as to generally have 'a desire to retain the desires that

are central to one's identity, a desire not to have one's psychology changed artificially, and a desire to live a meaningful life' (2022, 446). It may be that these desires are usually hidden or purely dispositional (Yu 2022, 452). We have already seen that desires need not be occurrent in order to be relevant to well-being (§2.4.1). If this argument is right, then we do not need to appeal to non-prudential reasons to explain why we tend to not have an all-things-considered reason to engineer our desires to better match the world. This is because, while doing so would improve some aspects of our well-being, the benefit of doing so would be outweighed by the countervailing harms incurred.

Nevertheless, Yu's approach does not solve the problem that motivates it. After all, sometimes the harm of frustrating the desires that she specifies would be outweighed by the benefit of engineering our desires to better fit the world. Yet, it is rare for one to introspectively identify the existence of any prudential reason whatsoever to acquire new desires simply to fulfil them. Yu's position does not capture this intuition. Moreover, many opportunities that we have to engineer our desires do not frustrate those desires that she stipulates. The previous examples of peripheral hobbies and leisure activities also serve to undermine Yu's position. In these cases, there are no countervailing prudential reasons that militate against engineering our desires to improve our well-being.

Therefore, the argument that human psychology is structured in such a way as to have hidden desires does not pacify this problem. At most, this argument serves only to partially blunt the force of the objection that motivates it. To my mind, this is not sufficient to warrant accepting it. Consequently, we ought to accept that prudential reasons are more complicated than may initially appear to be the case. On this view, we do not have a prudential reason to avoid making ourselves worse off by acquiring new desires that we are able to fulfil. If this is right, then Tolstoy has no reason to attempt to recover from his severe depression.

3.6 Chapter summary

I have argued that desire theories of well-being can account for the harm of severe depression. There are two ways in which severe depression typically diminishes well-being that these theories can recognise. Firstly, it does so by suppressing the motivational force of some desires. This inevitably frustrates those desires that we are left unmotivated to fulfil. Secondly, it does so by preventing pleasure from emerging. This inevitably frustrates desires that specify pleasure in their content. Consequently, severe depression is a state laden with desire frustrations. Nevertheless, these arguments explain some residual cases of severe

depression less well. For some of these cases, I have argued that self-destructive desires to be badly off can account for the harm of severe depression. For others, I have appealed to a deprivation account to explain how severe depression makes us worse off than we otherwise would have been. Often these aspects of depression's harm exist to various extents in combination with each other. These four complimentary explanations solve the problem of depression for desire theories of well-being and enrich our understanding of the psychological effects of depression.

Chapter 4 The Problem of Unstable Desires

4.0 Introduction

This chapter considers how desire theories of well-being ought to explain the relationship between the strength of a desire and the extent of the effects of its fulfilment or frustration on well-being. The second feature of the minimally plausible desire theory of well-being claims that the extent to which a subject is made better or worse off by a desire fulfilment or frustration is proportional to the strength of their fulfilled or frustrated desire (§1.2). However, this underspecifies the time at which the strength of a desire is relevant to its effects on well-being. According to one way of specifying this feature, simple concurrentism, it is the strength of a desire at the time of its fulfilment or frustration that determines the extent of these effects. The problem of unstable desires emerges when we observe that fleeting desires can be strong, long-standing desires can be weak, and desires in general are susceptible to fluctuations in strength. Consequently, simple concurrentism implies that you benefit more from the fulfilment of fleeting desires than is intuitive; that you benefit less from the fulfilment of long-standing desires than is intuitive; and, more generally, that the degree to which you benefit from the fulfilment of a desire varies, implausibly, with fluctuations in its strength. It has symmetrical implications about the extent of the harms incurred by frustrating desires. For these reasons, I argue that we should reject this view. I then introduce an alternative formulation of the theory that avoids this problem. Stability-adjusted desire theories of well-being claim that the average strength of a desire, and the length of time that it is held, both influence the extent to which its fulfilment or frustration affects well-being. I find that, rather than undermining desire theories of well-being, reflection on the problem of unstable desires illuminates a way in which these theories can better capture common intuitions about well-being. Consequently, insofar as one finds desire theories of well-being attractive, stability-adjusted versions of the theory are an improvement on how these views are usually constructed.

This chapter has the following structure: §4.1 outlines simple concurrentism and the problem of unstable desires. §4.2 argues that simple concurrentism cannot avoid this problem. §4.3 outlines stability-adjusted desire theories of well-being and defends these views from two objections. §4.4 considers value-fulfilment theories of well-being and argues that accepting this type of view is a less attractive solution to the problem of unstable desires. §4.5 considers

the implications of this argument for our understanding of prudential reasons and certain mental disorders. §4.6 concludes with a summary.

4.1 The problem of unstable desires

Fleeting desires are sometimes strongly held. This often seems to be the case when they coexist with strong emotions. For instance, when experiencing road rage, we may have a strong desire to shout at other motorists. Or, when elated, we may have a strong desire to give away significant amounts of money. Despite the strength of these desires, they are quickly lost. Other desires persist over time but are only moderately held. Take, for instance, the desire to finish writing a book or the desire to perfect a second language. In some moments these desires are strongly held, while at other times their strength wanes.

Some versions of the desire theory of well-being entail that fleeting desires, long-standing desires, and fluctuations in desire strength have counterintuitive effects on well-being. This is true of simple concurrentism. This view accepts the first feature of the minimally plausible desire theory (§1.2). However, it specifies the second in the following way:

> 2*. The extent to which a subject is made better or worse off by a desire fulfilment or frustration is proportional to the strength of their desire at the time of its fulfilment or frustration.

Whereas the proposed minimally plausible desire theory of well-being underspecifies how the strength of desire affects well-being, simple concurrentism claims that it is the strength of desire at the time of its fulfilment or frustration that determines the extent of these effects.

I focus on simple concurrentism for three reasons. Firstly, views of this sort are influential in the literature (Heathwood 2005, 490; Hare 1981, 103).[46] Secondly, whereas often desire theories of well-being underspecify how the strength of a desire affects well-being, simple concurrentism provides a precise account of this relationship. This clarity serves as a fruitful

[46] The term 'concurrentism' has been used to address several distinct questions in the literature. Here are four questions that can have a form of concurrentism as an answer: How must a desire be related to its object to count as fulfilled? How must a desire be temporally related to its owner in order to affect their well-being? At what time does the fulfilment or frustration of a subject's desire affect their well-being? How does the strength of a subject's desire determine the extent of its effects on their well-being? This chapter is concerned with this last question. §5.2.1 considers the wider question of whether desires need to exist at the same time as their object in order to improve well-being.

basis for discussion. Thirdly, framing this discussion around simple concurrentism serves as a useful dialectical device to highlight how desire theories of well-being more generally can avoid the problem of unstable desires. As we will see in §4.3, findings from this discussion are also relevant to versions of the desire theory of well-being that reject the claim that desires must exist at the same time as their objects in order to improve well-being. Consequently, this discussion has implications for desire theories of well-being more broadly.

Simple concurrentism shares the attractions of the minimally plausible desire theory of well-being outlined in §1.2. These are: its ability to capture the resonance requirement, its ability to be conceptually parsimonious, its ability to postulate a universal standard of well-being, its operationalisability in policymaking, and its compatibility with liberal political philosophy. Additionally, simple concurrentism has two further attractions. Firstly, this view can put forward its own proposed solution to the problem of Dead Sea apples (§5.2.1). Dead Sea apples describe situations where we desire something, get it, and find ourselves disappointed and devoid of feelings of satisfaction (Sidgwick 1907, 110). They are alleged counterexamples to the view that fulfilling desires always improves well-being. Simple concurrentism can claim that, in these cases, the desire ceases immediately prior to the conditions of its fulfilment occurring. If this is right, then the desire does not improve well-being (Heathwood 2005, 493).[47] Secondly, this view can avoid the arguably counterintuitive conclusion that fulfilling lost desires improves well-being (Bykvist 2003, 116). While there is considerable debate over whether some lost desires remain relevant to well-being, the consensus is that at least some are not. Because simple concurrentism requires that desires temporally overlap with the conditions of their fulfilment in order to improve well-being, this theory provides a clear explanation of why lost desires do not improve well-being.

If simple concurrentism is correct, then strongly held fleeting desires matter more to well-being than moderately held long-standing desires. For example, this view entails that the fulfilment of the strong fleeting desire to shout at another motorist when experiencing road rage improves well-being more than that of the moderate long-standing desire to finish writing a book. Moreover, on this view, fluctuations in desire strength have counterintuitively large effects on well-being. For instance, if the strength of a desire to finish

[47] William Lauinger argues that this approach fails because there remain cases where we are desiring, getting, but not benefitting from the object of our desire (2011, 311). If he is right, then one of the attractions of simple concurrentism is revealed as illusory. In §5.2.1 I argue that this proposed solution to the problem of Dead Sea apples fails.

writing a book decreases in the moments preceding its fulfilment, then its fulfilment improves well-being substantially less than had it been fulfilled mere moments beforehand. These are troubling conclusions.

Part of the appeal of desire theories of well-being is that they tie well-being to the subjective attitudes that we most closely identify with. However, the long-standing desires that we typically consider to be central to our identity are often prone to fluctuations in strength. They are therefore systematically misvalued by simple concurrentism. This problem is compounded when we consider that we can feel alienated from desires that conflict with our values (Hubin 2003, 326–328). Alienation seems to be more commonly experienced in relation to fleeting desires. Take, for instance, the strong but short-lived desires of a recovering addict to return to drug use or gambling (Lewis 2000, 70). Because simple concurrentism overvalues the effects that fulfilling and frustrating fleeting desires have on well-being, it places too much importance on those subjective attitudes that we consider peripheral or contrary to our identity.

These concerns form what can be called 'the problem of unstable desires'. It can be summarised as follows:

P1: Viable theories of well-being must value the effects of fleeting desires, long-standing desires, and fluctuations in desire strength on well-being in an intuitive way.

P2: Simple concurrentism does not value the effects of fleeting desires, long-standing desires, and fluctuations in desire strength on well-being in an intuitive way.

C: Therefore, simple concurrentism is not a viable theory of well-being.

This argument is valid. Moreover, P1 is compelling. In the next section I argue that simple concurrentism cannot avoid P2. Consequently, I find that we ought to reject this view. In §4.3 I put forward a version of the desire theory of well-being that avoids this problem.

4.2 Simple concurrentism

Common intuitions suggest that the fulfilment and frustration of long-standing desires generally affects well-being more than that of fleeting desires. Proponents of simple concurrentism can attempt to capture this intuition by appealing to a distinction between phenomenological and overall strength of desire. According to desire theories of well-being, it is the overall strength of the desire, rather than that of its phenomenology, that determines

the extent of its effects on well-being (Griffin 1986, 14–15). If the two are distinct, then proponents of the view can make the following three claims:

1. While fleeting desires may sometimes feel strong, they are normally weakly held.

2. While long-standing desires may sometimes feel weak, they are normally strongly held.

3. While we may sometimes feel fluctuations in desire strength, desires are normally relatively stable.

This approach distinguishes between the overall strength of desire and the strength of its phenomenology. There are independent reasons to think that this distinction is necessary to explain our mental lives. Call the view that the overall strength of desire is proportional to the strength of its phenomenology 'proportionalism about desire and phenomenology'. This is different from the version of proportionalism discussed in Chapters Two and Three. That view is about the relationship between strength of desire and strength of motivation. The view discussed in this section is about the relationship between strength of desire and strength of phenomenology.

Proportionalism about desire and phenomenology is not an attractive view. Some of the reasons why we ought to reject this view overlap with the reasons why we ought to reject proportionalism about desire and motivation (§2.3). For instance, we are better able to explain the psychological phenomena discussed in §2.4.1 by rejecting this version of proportionalism as well. Weakness of will, forgotten desires, and unusually strongly felt desires often involve desires that have a disproportional phenomenological strength from their overall strength. For instance, forgotten desires have no phenomenology. Yet, it is counterintuitive to suggest that these desires no longer exist when our attention is not focussed on them. Moreover, when experiencing weakness of will, the phenomenological effects of desires are often far stronger than their overall strength. Consider, for instance, the abstinent smoker capitulating to the fleeting desire for a cigarette. One intuitive explanation of this case is that the phenomenology of the desire for the cigarette focusses attention and amplifies the motivational effects of the desire. Unusually strongly felt desires also contribute to this picture. Consider again the desires that emerge during road rage. While many people strongly feel these desires, far fewer reflectively endorse them as having been strongly held.

Most damagingly, proportionalism about desire and phenomenology entails that we infrequently desire to retain the things that we already have.[48] Desires of this sort often lack a phenomenology. Nevertheless, it is counterintuitive to claim that we have no such desires. Take, for instance, the desire to remain employed. While its phenomenology can be elicited through prompting, it is usually dormant until we focus our attention or perceive that our employment is threatened. An understanding of desire that identifies overall strength of desire with strength of its phenomenology is unable to explain our strong but phenomenologically dormant desires to retain the things that we already have. Consequently, we have independent reasons to reject this view.

However, this approach cannot completely defuse the problem of unstable desires. There is nothing in this picture that disbars fluctuations in desire strength from having counterintuitively large effects on well-being. To prevent this from happening we would need to postulate a degree of desire stability that far exceeds common intuitions. There appears to be no independent reasons to hold such a view. Moreover, rejecting proportionalism about desire and phenomenology does not prevent the fulfilment and frustration of fleeting desires from having counterintuitively large—and long-standing desires from having counterintuitively small—effects on well-being. This argument can only plausibly claim that such occurrences are less frequent than our phenomenology suggests. Therefore, rejecting this view only succeeds in partially diminishing the force of the problem. It fails to solve it.

Even if we are content to live with a diminished problem of unstable desires, there is an additional issue that simple concurrentism faces. Chris Heathwood identifies this:

> ‘What if the intensity of the desire changes over the time that it is being concurrently satisfied? We can avoid this by defining desires as things that occur at instants (or at very brief intervals of time). At each brief interval at which some concurrent desire satisfaction occurs, there is just one intensity. We can say that the intensity of a desire has ‘changed’ when a person has a desire of some intensity for some proposition at

[48] Some readers may be concerned that this argument is incompatible with the death of desire principle. This is the claim that desires always cease to exist once we perceive their fulfilment (Pineda-Oliva 2021, 247). However, my objection is based on the failure of proportionalism about desire and phenomenology to account for our desires to retain the things that we already have. These are future-directed desires. Consequently, the objection stands even if the death of desire principle is correct.

> some brief interval and then, at the next brief interval, has a desire of a different intensity for the same proposition' (Heathwood 2005, 490n5).

Here Heathwood addresses the question of how simple concurrentism ought to account for fluctuations in desire strength that happen during the time of fulfilment and frustration. Simple concurrentism claims that the extent to which well-being is affected is determined by the strength of desire at the time of its fulfilment or frustration. However, it is unclear how the view handles cases where there are multiple levels of desire strength during that time. For instance, consider the desire to ride a rollercoaster. This may be strong at the start of the ride, weak during a particularly steep incline, and moderate in the moments preceding the ride's end. This desire has three different levels of strength during the time of its fulfilment. Simple concurrentism should have something to say about such cases.

Heathwood responds to this problem by claiming that there is a single level of desire strength that covers instants or brief intervals of time. If this is right, then examples like the rollercoaster ride are composed of a series of consecutive desire fulfilments about the same thing, rather than fluctuations in desire strength during the time of fulfilment. Call this the desire-at-instant view.

The desire-at-instant view can explain how apparent fluctuations in desire strength during the time of fulfilment and frustration affect well-being. It does so by denying that such occurrences involve actual fluctuations in desire strength. Instead, this view claims that the appearance of such fluctuations is explained by a series of consecutive desires about the same thing. Nevertheless, this approach requires the revision of common intuitions about desires. Our phenomenology and everyday language both point towards desires existing for longer periods of time. Consider your long-term desires to remain in a close relationship with a partner, friend, or family member. Desires of this sort are intuitively better captured by a theory that allows them to exist for more than a brief interval or instant. Moreover, as previously noted, the persistence of desire over time is part of our folk concepts of personal identity (§2.4.1). Therefore, the desire-at-instant view forces us to revise these concepts. Of course, these intuitions and folk concepts may be worth revising if there are independent reasons for accepting the desire-at-instant view. However, to my knowledge, no such reasons exist. Consequently, this approach to salvaging simple concurrentism should be rejected.

Moreover, this view also has counterintuitive implications for well-being. This is because it significantly overvalues the well-being effects of desires for processes and undervalues those of desires for outcomes. Examples of desires for processes might include the desire to play a game of football, watch a film, or run a half marathon. Conversely, examples of desires for

outcomes might include the desire to finish writing a book, reach the summit of Mt. Everest, or have nations sign up to an international emissions reduction treaty. Of course, it is possible that these latter examples could also involve desires for processes as well. However, let us set aside that possibility for illustrative purposes. It is true that there are at least some desires that are entirely concerned with processes and at least some others that are entirely concerned with outcomes.

Assume that my examples of desires for outcomes exist alongside their fulfilment conditions for only an instant. The moment the last word is penned, the summit is reached, or the treaty is signed, is the only time that the desire exists alongside its fulfilment conditions. Conversely, imagine that the desires for processes listed exist for far longer alongside their fulfilment conditions. Assume that the instants that Heathwood has in mind are one second long and that the processes that I have listed last for ninety minutes. That means that these processes involve 5,400 consecutive desire fulfilments. Conversely, the outcomes listed involve only a single desire fulfilment. Consequently, these desires for outcomes would need to be far stronger than the desires for processes to have the same effects on well-being. Clearly this is an unacceptable result. Desires for processes do not matter overwhelmingly more to well-being than desires for outcomes. Heathwood's view fails to capture this intuition.

A possible response to this problem claims that the benefits of fulfilling desires for outcomes primarily derive from our backwards-facing desires about them. For instance, while fulfilling the desire to reach the summit of Mt. Everest only marginally increases well-being at the time of reaching its peak, the fulfilment of backwards-facing desires about this accomplishment continues to improve well-being long after the event. If this is right, then the gap between our valuations of outcomes relative to processes may narrow. Nevertheless, this approach is unviable in at least some cases. Consider the following counterfactual: The world ends after you fulfil your desire to play a ninety-minute football game, or the world ends after you fulfil your desire to complete your life goal of reaching the summit of Mt. Everest. In this case, it is not possible to appeal to backwards-facing desires to explain why the well-being effects of the former do not drastically outweigh those of the latter. Even in less dramatic scenarios, this approach does not work. This is especially true of long-standing desires fulfilled later in life. In these cases, we may not have much time remaining to have backward-facing desires about these outcomes. Yet, many of our lifetime goals are only achieved in later life. Consequently, our theory of well-being should not rely on backwards-facing desires to explain why these goals are important to our lifetime well-being.

Simple concurrentism can partially diminish the force of the problem of unstable desires by rejecting proportionalism about desire and phenomenology. Nevertheless, even if we are content to live with a diminished problem of unstable desires, this view struggles to answer a related problem. This is about how simple concurrentism ought to account for fluctuations in desire strength during the time of fulfilment or frustration. The only candidate explanation of this in the literature, an appeal to the desire-at-instant view, fails on this account. For these reasons we should reject simple concurrentism and look for an alternative theory of well-being.

4.3 Stability-adjusted desire theories of well-being

We can construct a version of the desire theory of well-being that captures many of the attractions of simple concurrentism and is not rendered implausible by the problem of unstable desires. Moreover, this view can explain how fluctuations in desire strength during the time of fulfilment or frustration affect well-being. It does so by rejecting the claim that the strength of a desire at the time of its fulfilment or frustration is what determines the extent of its effects on well-being. Instead, it calculates these effects by appealing to two principles. The Averaging Principle claims that the lifetime average strength of a desire governs the extent of its effects on well-being, while the Longevity Principle claims that the length of time that a desire is held magnifies these effects.[49] Call this view the 'stability-adjusted desire theory of well-being'. It accepts the first feature of the proposed minimally plausible desire theory of well-being (§1.2), but specifies the second in the following way:

> 2**. The extent to which a subject is made better or worse off by a desire fulfilment or frustration is proportional to the lifetime average strength of their fulfilled or frustrated desire and magnified by the length of time that the desire is held.

[49] There are precedents for the view that the longevity of a desire magnifies the extent of its effects on well-being. For instance, Krister Bykvist tentatively defends the view that the moral significance of a preference is amplified by its longevity (2003, 128). My view differs from Bykvist's in two important respects. Firstly, whereas Bykvist claims that desires must exist at the same time as their objects in order to be significant, I remain neutral on this question. Secondly, unlike Bykvist, my view claims that the average lifetime strength of a desire also influences the extent of its significance. Additionally, my paper outlines a distinct motivation for accepting the longevity principle and considers a separate counterargument to it. Heathwood also notes that some versions of the desire theory of well-being claim that the longevity of a desire magnifies the extent of the effects of its fulfilment or frustration on well-being (2015, 135).

The Averaging Principle prevents fluctuations in desire strength from having counterintuitively large effects on well-being, while the Longevity Principle entails that long-standing desires generally affect well-being more than fleeting desires. Consequently, this view better captures intuitions about the effects of fleeting desires, long-standing desires, and fluctuations in desire strength on well-being. This avoids the problem of unstable desires that weakened the appeal of simple concurrentism. Moreover, this view can explain how fluctuations in the strength of desire during the time of its fulfilment or frustration affect well-being. It does so by applying the Averaging and Longevity Principles to the whole time that a desire is held. This includes the time at which a desire is held concurrently alongside the conditions of its fulfilment or frustration. Consequently, there is no need to appeal to the desire-at-instant view.

Unlike simple concurrentism, the stability-adjusted desire theory of well-being does not entail that desires must exist at the same time as their objects in order to improve well-being. Consequently, this view lacks some of the attractions of simple concurrentism. For instance, it does not entail that fulfilling lost desires has no effect on well-being, and it cannot put forward the same proposed solution to the problem of Dead Sea apples that simple concurrentism can (§4.1).[50] Nevertheless, this view retains most of the more general benefits of desire theories of well-being (§1.2). For instance, it shares their ability to capture the resonance requirement, their ability to postulate a universal standard of well-being, their operationalisability in policymaking, and their compatibility with liberal political philosophy. If we want this theory to capture the additional attractions of simple concurrentism, then we can modify the stability-adjusted desire theory of well-being by postulating an additional feature:

3. Only desires that exist concurrently with their objects non-instrumentally improve well-being.

The resultant version of the desire theory of well-being accepts features 1, 2**, and 3. Call this view ‘stability-adjusted concurrentism’. Whereas simple concurrentism entailed 3 through its specification of 2*, stability-adjusted concurrentism must explicitly adopt this

[50] There are compensatory advantages to rejecting concurrentism. For instance, some writers argue that some lost desires do improve well-being if their objects occur after the desire ceases (Bruckner 2013; Dorsey 2013; Vorobej 1998). One advantage of this approach is that it can capture some intuitions about the possibility of posthumous benefits and harms (Pitcher 1984). The debate between concurrentism and non-concurrentism is discussed in more detail in §5.2.1. I limit myself here to the claim that both versions are made more plausible by adopting 2**.

feature. In this thesis, I remain neutral between the stability-adjusted desire theory of well-being and stability-adjusted concurrentism. The discussion that follows applies equally well to both views. For simplicity, I will simply refer to stability-adjusted desire theories of well-being going forward. I am using this as an umbrella term for all views that accept 2**.

Stability-adjusted desire theories of well-being sacrifice some of the conceptual parsimoniousness that makes desire theories of well-being initially attractive. Notably, they must explain how to balance desire longevity with desire strength. Probably the simplest way of doing so involves identifying units of desire strength and units of temporal length and simply multiplying them together. Furthermore, if we take this approach, then we ought to factor in a diminishing rate of added value for each additional temporal unit prior to multiplication. This prevents the theory from counterintuitively entailing that weakly held long-standing desires matter overwhelmingly more to well-being than much stronger fleeting desires. If we want to fully guard against the possibility of such conclusions, then we may want the rate to decrease to zero eventually. Factoring in a diminishing rate of added value for each additional temporal unit means that fulfilling long-standing desires improves well-being progressively more than fulfilling fleeting desires, but not to the extent that fleeting desires become largely irrelevant to well-being. This strikes me as an intuitive result. Nevertheless, postulating a diminishing rate of added value for each additional temporal unit does further compromise the extent to which these theories can remain conceptually simple.

Adopting 2** can solve the problem of unstable desires. Nevertheless, some readers may be concerned that this solution is intolerably arbitrary. Without independent reason for this way of specifying 2, then the adoption of 2** could be perceived as an arbitrary manoeuvre solely designed to evade a narrow set of counterexamples. However, I do not think that the charge of arbitrariness is warranted. There are two reasons for this. Firstly, the way that desire theories of well-being are usually constructed underspecifies how exactly strength of desire affects well-being. These theories need to answer this question. Simple concurrentism is no less arbitrary than stability-adjusted desire theories of well-being in doing so. Moreover, as we have seen, simple concurrentism is unable to explain fluctuations in desire strength during the time of fulfilment or frustration. Therefore, stability-adjusted desire theories of well-being may well be the simplest versions of the desire theory of well-being that can explain this phenomenon. Secondly, the move from 2 to 2** is not solely motivated by a narrow set of counterexamples. Rather, this specification is made to better capture common intuitions about well-being in a range of cases. I pointed out earlier that part of the appeal of desire theories of well-being is that they tie well-being to the subjective attitudes that we

most closely identify with (§4.1). Although these theories do not take proximity to identity to non-instrumentally magnify the effects of a desire's fulfilment or frustration on well-being, they are made more plausible by entailing that the desires that we identify most with generally affect well-being more than other desires.[51] Consequently, there is a principled reason for preferring this formulation of the view.

In addition to facing a larger explanatory burden, stability-adjusted desire theories of well-being face two extra problems. Firstly, these views entail that, other things being equal, deliberately delaying the fulfilment of desires increases the amount of well-being that their fulfilment produces. This is because the Longevity Principle claims that the length of time that a desire is held prior to its fulfilment or frustration magnifies the extent of its effects on well-being. For instance, if I can fulfil my desire to go for a run at any point during the day, then, other things being equal, these views entail that deliberately delaying doing so until the evening is the most prudent choice. This can seem unintuitive. Yet, if the Longevity Principle is correct, then the conclusion follows.

While this conclusion may seem unintuitive, it is not unpalatable in all cases. Sometimes we have desires that specify a wide time preference in their content. This happens when our desires specify multiple possible times at which their object can occur. It strikes me that deliberately delaying the fulfilment of desires of this sort may well increase the amount of well-being produced by their fulfilment. Take, for instance, the desire to climb Mt. Everest at any time between now and one's death. Over time, what started as a flight of fancy may become a serious life goal that increasingly becomes part of one's sense of identity. Consequently, it does not strike me as intolerably unintuitive to accept that deliberately delaying the fulfilment of this type of desire amplifies the amount of well-being produced by its fulfilment. Therefore, we ought not to rule out the idea that well-being is improved by deliberately delaying the fulfilment of some desires. However, this is an unconvincing response for the bulk of cases. There are many desires that we aim to fulfil as quickly as possible. Take, for instance, the desire to drink water when thirsty. It is counterintuitive to claim that deliberately delaying the fulfilment of this desire increases the amount of well-being that its fulfilment produces.

[51] Some versions of the desire theory of well-being do require a certain level of identification with desires in order for them to be relevant to well-being (Noggle 1999, 314–316). The arguments of this thesis do not rely on restricting which desires affect well-being in this way.

In response to cases like this, we can claim that the non-instrumental benefit of deliberately delaying the fulfilment of a desire is sometimes outweighed by the additional harms incurred by doing so. There are two ways in which this seems to be true. Firstly, deliberately delaying the fulfilment of a desire sometimes means frustrating other desires. Consider again the example of deliberately delaying the fulfilment of the desire to drink water when thirsty. According to stability-adjusted desire theories of well-being, this non-instrumentally increases the amount of well-being that its fulfilment produces. Nevertheless, the delay also frustrates other desires. For instance, it frustrates the standing desire to be free from aversive mental states (§1.3; §3.4). Consequently, the benefits of deliberately delaying the fulfilment of desires are sometimes outweighed by the acquisition of additional harms. Therefore, stability-adjusted desire theories of well-being do not entail that it is always prudent to deliberately delay the fulfilment of desires.

There is a second way that these views can recognise that deliberately delaying the fulfilment of desires does not always improve well-being. This involves making a distinction between unfulfilled desires that are in a state of frustration and those that are in a state of anticipation.[52] Sometimes our well-being is decreased by the absence of the things we desire.[53] This happens when our desire aims at immediate fulfilment. Take, for instance, a child's desire to open presents before Christmas Day. This desire is in a state of frustration for the length of time that it is held prior to its fulfilment. While prolonging the length of time that this desire is held does non-instrumentally increase the amount of well-being produced by its fulfilment, doing so also non-instrumentally decreases well-being by prolonging its frustration prior to that fulfilment. Therefore, delaying the fulfilment of desires that aim at immediate fulfilment does not improve aggregate well-being.

---

[52] Kris McDaniel and Ben Bradley argue that there are additional states that desires can be in aside from fulfilment or frustration. They argue that desires can also be cancelled (McDaniel & Bradley 2008, 274). This happens 'when a person's desire that P is conditional on Q, the desire that P is cancelled if and only if Q is false' (2008, 275). On their view, conditional desires with unsatisfied conditions are cancelled, rather than fulfilled. The idea of desire cancellation as an alternative state to fulfilment or frustration is a precedent for my claim that desires which specify a wide time preference in their content are not frustrated until the specified window of time has elapsed.

[53] Daniel Pallies has recently argued that the concept of desire is divisible into attractions and aversions. He argues that when we have an aversion to something, avoiding that thing does not non-instrumentally improve well-being. Conversely, when we have an attraction to something, failing to get that thing does not diminish well-being (Pallies 2022). This distinction allows him to explain how some desires operate purely positively or negatively on well-being. If this is right, then there are additional resources available to desire theories of well-being to explain why delaying the fulfilment of desires with a wide time preference does not diminish well-being.

Conversely, some desires are in a state of anticipation before fulfilment. Take, for instance, the desire to go for a walk at any time between now and sunset. This desire specifies a wide time preference. Consequently, it is not frustrated until the end of the timeframe that it specifies. In cases like this, stability-adjusted desire theories of well-being entail that deliberately delaying the fulfilment of the desire increases the well-being produced by its fulfilment without incurring countervailing harms. If this is right, then stability-adjusted desire theories of well-being can claim that deliberately delaying the fulfilment of some desires does improve well-being, while deliberately delaying the fulfilment of others does not. The way to distinguish between which desires have which effects on well-being is dependent upon whether the desire specifies a wide time preference or aims at immediate fulfilment. Consequently, stability-adjusted desire theories of well-being have two ways of explaining why deliberately delaying the fulfilment of desires does not always improve well-being, and one argument in favour of the conclusion that sometimes deliberately delaying the fulfilment of a desire does magnify the amount of well-being that its fulfilment produces.

There is a second problem worth considering for stability-adjusted desire theories of well-being. This concerns the intuition that the strength of desire at the time of its fulfilment or frustration matters at least slightly more to well-being than the strength of desire at other times. Adopting the Averaging Principle means that the theory struggles to capture this intuition. To illustrate the intuitive force of this objection, consider the following two desires:

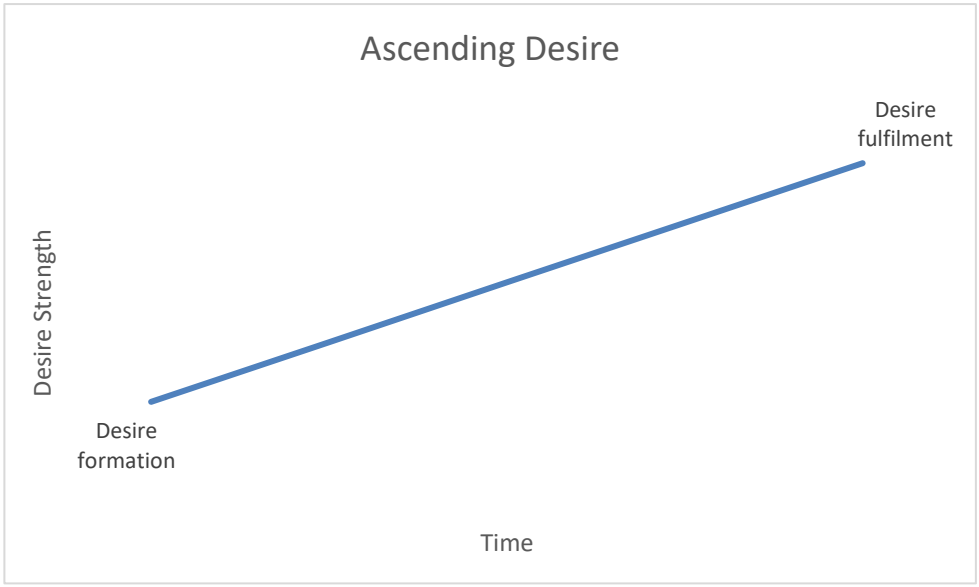

Figure 1. Ascending Desire

Ascending Desire starts as a relatively weak desire. Over time it gradually becomes stronger. At the time of its fulfilment, the desire is three times stronger than when it was formed.

Consider a second desire:

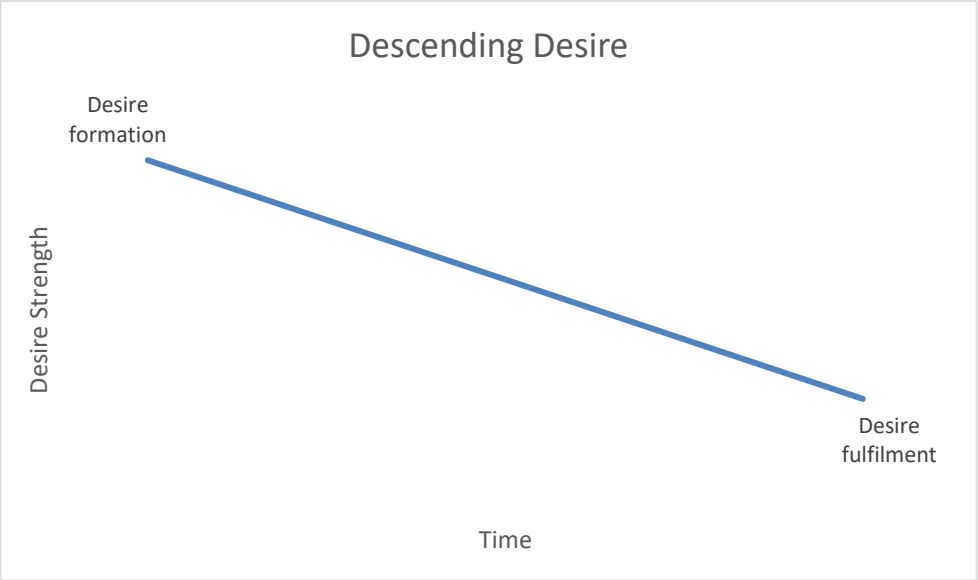

Figure 2. Descending Desire

Descending Desire has the inverse shape to Ascending Desire. It starts strong. Over time it gradually becomes weaker. At the time of its fulfilment, the desire is three times weaker than when it was formed.

Stability-adjusted desire theories of well-being imply that these desires have identical effects on well-being. This is because both desires are held for the same length of time and have the same average strength. Nevertheless, there is some intuitive force in the idea that the fulfilment of Ascending Desire increases well-being at least slightly more than that of Descending Desire. Stability-adjusted desire theories of well-being ought to have something to say about this intuition.

It is worth pointing out that not everyone shares the intuition that Ascending Desire matters more to well-being than Descending Desire. Indeed, this intuition appears to be based on the observation that fulfilling Ascending Desire often feels better than fulfilling Descending Desire. However, as we have seen, desire theories of well-being are already committed to rejecting the idea that feelings determine well-being (§1.2). Indeed, this is one of the attractions of these theories. It is what allows them to claim that pleasures generated by misperception and deception do not benefit us as much as those generated in response to veridical perceptions and true beliefs. Consequently, those who find desire theories of well-being attractive are unlikely to be greatly troubled by the fact that these theories fail to capture the intuition that Ascending Desire matters more to well-being than Descending Desire. Nevertheless, the fact that fulfilling Ascending Desire often feels better may reveal one way in which it is more beneficial to us. It seems likely that we are the sort of creatures that have standing desires for feelings or attitudes such as pleasure or satisfaction (§1.3). If

this is right, then fulfilling Ascending Desire often improves well-being more than fulfilling Descending Desire because it better fulfils those standing desires. Consequently, stability-adjusted desire theories of well-being can capture the intuition that Ascending Desire often improves well-being more than Descending Desire.[54]

We have seen that stability-adjusted desire theories of well-being can navigate the problem of unstable desires and avoid the counterintuitive conclusions of simple concurrentism. They do so at the expense of acquiring the additional complexity of needing to explain how to balance desire strength with desire longevity. Moreover, these views are subject to two additional objections. The first one states that they counterintuitively entail that deliberately delaying the fulfilment of desires improves well-being. The second one claims that these theories are unable to capture the intuition that Ascending Desire matters more to well-being than Descending Desire. I have shown that stability-adjusted desire theories of well-being can capture intuitions in both cases.

4.4 Value-fulfilment theories of well-being

All good theories risk being supplanted by better ones. While stability-adjusted desire theories of well-being can avoid the counterintuitive conclusions that undermined simple concurrentism, their increased complexity may make them less attractive than rival theories of well-being. Some of the concerns that motivate the adoption of stability-adjusted desire theories of well-being are appealed to in support of value-fulfilment theories of well-being (henceforth value-fulfilment theories).

The word 'values' can be suggestive of complex normative commitments. However, proponents of value-fulfilment theories often have in mind a wider conception of values than this. They consider values to be a subset of our pro-attitudes. That subset is restricted to pro-

[54] This discussion resembles a debate about whether the shape of a life non-instrumentally affects a subject's lifetime well-being. David Velleman argues that lives with an upward well-being trajectory are, other things being equal, non-instrumentally better for the subject living them than lives with a downwards well-being trajectory (1991). Conversely, Fred Feldman argues that the shape of a life does not non-instrumentally affect the subject's lifetime well-being (2004, ch.6). He points out that lives with an upward trajectory may be instrumentally better because many people prefer such lives. There are some parallels between Feldman's position on this debate and my analysis of Ascending and Descending Desire. Nevertheless, intuitions about the importance of life trajectory do not necessarily translate into similar intuitions about whether Ascending Desire matters more than Descending Desire. The discussions may be parallel, but the issues are distinct.

attitudes that conform to a specified set of requirements. Normally those requirements are largely concerned with how the pro-attitude relates to other aspects of our mental lives (Raibley 2010, 608). This allows us to conceive of values as a relatively broad set of pro-attitudes. These need not be as cognitively sophisticated as the word 'value' sometimes implies (Yelle 2016, 1416). Consequently, these views should not be immediately dismissed for excluding too much of what we intuitively think non-instrumentally affects well-being.

A compelling version of the value-fulfilment theory is put forward by Jason Raibley. He argues that a subject's well-being is primarily determined by the realisation of their values, and the cultivation of dispositions towards the realisation of those values (Raibley 2010, 596). He defines valuing as the possession of a pro-attitude towards a state of affairs that the subject stably identifies with (Raibley 2010, 606–607). The extent to which the realisation of a value non-instrumentally improves well-being is dependent upon the pro-attitude's intensity and the extent to which the subject identifies with it (Raibley 2010, 608). We may want to add to this view the requirement that values must be autonomously acquired in order for their realisation to improve well-being (Yalle 2014). This prevents artificially aroused pro-attitudes from affecting well-being.

Raibley thinks that the pro-attitudes that we identify with normally have at least three components. Firstly, they have a phenomenology that is distinct from that of pro-attitudes alienated from our identity. Secondly, we are disposed towards approving of the pro-attitude and we take it as representative for who we want to be. Thirdly, we perceive the pro-attitude as reason-giving in the sense of justifying actions. It is the realisation of pro-attitudes that conform to these criteria that primarily improves well-being. Raibley supplements his view by also noting that some affective states also non-instrumentally affect well-being (2010, 609).

Part of the appeal of value-fulfilment theories stems from the observation that our desires are not always reflective of our deepest concerns (Raibley 2010, 599; Yelle 2014, 372). Some desires are alienated from our identity despite resonating with other aspects of our psychology. This is often true of our fleeting desires. According to value-fulfilment theories, the fact that we are less likely to stably identify with these desires means that their fulfilment often does not benefit us (Raibley 2010, 614; Yelle 2014, 374). Consequently, these views are well-placed to avoid some of the counterintuitive conclusions of simple concurrentism. However, the subset of desires that are alienated from our identity is not limited to fleeting desires. Some people have the intuition that the fulfilment of long-standing desires that are alienated from our identity also does not benefit us. If we share this intuition, then value-

fulfilment theories have an advantage that stability-adjusted desire theories of well-being lack. They can account for unstable desires *and* address concerns about alienation.

If value-fulfilment theories can better capture intuitions in these cases, then their appeal may eclipse that of stability-adjusted desire theories of well-being. However, there are good reasons to think that value-fulfilment theories are the less convincing family of views. I focus on three arguments in favour of this conclusion. The first highlights the larger explanatory burden that they incur. The second undermines the reasons for finding these theories attractive in the first place. The third points out that a version of the problem of unstable desires is reproducible for value-fulfilment theories.

Firstly, value-fulfilment theories are more conceptually convoluted than desire theories of well-being. Not only must proponents of these theories construct a compelling theory of the valuing attitude, but they must also accept that well-being is constituted by multiple components. This is necessary to avoid the conclusion that beings that are incapable of valuing are also incapable of having well-being (Lin 2017a, 357–365). This would exclude babies, many non-human animals, and some adult humans from being subjects of well-being.[55] This is clearly an unacceptable result. It can be avoided by postulating additional things that non-instrumentally affect well-being. For instance, it may be that value-fulfilment and some affective states are what determines well-being (Raibley 2010, 609). However, adopting a pluralistic theory of well-being means that value-fulfilment theories sacrifice some conceptual parsimoniousness. Consequently, they incur a larger explanatory burden than that of stability-adjusted desire theories of well-being.

Secondly, the reasons for favouring value-fulfilment theories over desire theories of well-being are not particularly strong. If one is drawn to a value-fulfilment theory partly because of the problem of unstable desires, then stability-adjusted desire theories of well-being can avoid these troubling implications. Value-fulfilment theories and stability-adjusted desire theories of well-being differ on this issue because the former claims that many fleeting

[55] An alternative approach is to claim that value-fulfilment theories only account for the well-being of subjects with the capacity to value (Yelle 2016). It may be that other beings are subjected to different standards of well-being (§1.4). However, this approach inherits the unenviable task of needing to explain which standards apply to subjects that lack this capacity. Moreover, the view must accept that sometimes the standards of a subject's well-being change when they acquire or lose the capacity for valuing. This means that a subject's well-being could radically change for better or worse when they acquire or lose this capacity (Lin 2017a, 359–360). Conversely, an advantage of desire theories of well-being is that they can postulate a universal standard of well-being applicable to all beings capable of well-being (§1.2).

desires do not affect well-being, whereas the latter claims that these effects are simply less pronounced than those of long-standing desires. Nevertheless, even if one finds these cases more intuitively explained by value-fulfilment theories, it strikes me as unlikely that this benefit is worth the additional explanatory burden that these theories incur.

The main advantage that value-fulfilment theories have over stability-adjusted desire theories of well-being is that the former can exclude desires that we do not identify with from affecting our well-being.[56] Nevertheless, it is unclear to me whether we ought to exclude these desires. The strongest reasons for doing so appear to be in response to cases of addictions and compulsions. Take, for instance, desires born of drug addiction. It may strike us as unpalatable that the fulfilment of such desires benefits the addict. Indeed, it is intuitive to think that fulfilling addictive desires harms the addict because doing so sustains their underlying addiction. Yet, if desire theories of well-being are correct, then fulfilling desires born of addiction does improve the well-being of the addict. Consequently, one might opt for a value-fulfilment theory over a desire theory of well-being to avoid this conclusion.

However, while proponents of desire theories of well-being are forced to concede that fulfilling desires born of addiction is of some benefit to the addict, doing so may still be an all-things-considered harm (Heathwood 2005, 493–494). This is because the benefit of fulfilling desires born of addiction is outweighed by the countervailing harms incurred from doing so. These harms arise from the damage to health, finances, time, and relationships that addiction frequently results in. This frustrates more and stronger desires than the addiction typically fulfils. Consequently, we do not need to exclude addictive desires from non-instrumentally affecting well-being in order to explain why they are nevertheless bad for us.

It is possible to construct hypothetical examples whereby fulfilling a desire born of addiction does not lead to these countervailing harms, does not produce any desirable mental states in response to being fulfilled, but nevertheless does motivate us to continue to acquire the addictive experience or substance. Derek Parfit appeals to this sort of example in order to disprove versions of the desire theory of well-being that take all desires to be relevant to well-being:

[56] It is possible to formulate a desire theory of well-being that also restricts the desires that count towards well-being to only those that we identify with (Noggle 1999, 314–316). Depending on how we define valuing, views of this sort may blur the boundaries between value-fulfilment theories and desire theories of well-being.

> ‘I tell you that I am about to make your life go better. I shall inject you with an addictive drug. From now on, you will wake each morning with an extremely strong desire to have another injection of this drug. Having this desire will be in itself neither pleasant nor painful, but if the desire is not fulfilled within an hour it will then become very painful. This is no cause for concern, since I shall give you ample supplies of this drug. Every morning, you will be able at once to fulfil this desire. The injection, and its after-effects, would also be neither pleasant nor painful. You will spend the rest of your days as you do now’ (Parfit 1984, 497).

It would appear that desire theories of well-being are forced to concede that you are benefitted by this inflicted dependency on the drug. After all, your life contains more desire fulfilment than it would have done without the onset of addiction. Moreover, the example is designed so as to make it impossible to appeal to countervailing desire frustrations to explain the addiction as an all-things-considered harm.

I confess that it strikes me as counterintuitive to conclude that fulfilling this type of desire improves our overall well-being. Nevertheless, I do not think that arguments of this sort should be considered fatal to desire theories of well-being. We should be more sceptical of intuitions generated by extreme thought experiments than we are of real-world cases. Desire theories of well-being get the intuitive answer in real-world cases of addiction. We should not require viable theories of well-being to generate intuitive conclusions in every extreme hypothetical scenario as well. Moreover, the fact that this type of addiction may improve our overall well-being does not mean that we have any prudential reason to acquire an addiction of this sort. We do not have prudential reasons to generate new desires simply so that they can be fulfilled (§3.5).

Furthermore, addictive desires of the sort described in Parfit’s example may still qualify as an all-things-considered harm providing that we have a second-order desire to be free from them. It strikes me that most people in this situation would develop such desires. It is not particularly common to want to be the sort of person who is dependent upon a substance that has neither affective nor instrumental benefits. Judgements of this sort are perhaps what undergird the intuition that we would not be benefitted by addictions of this sort. After all, the way the thought experiment is constructed suggests that we get nothing out of fulfilling these desires. Consequently, it would be unusual to want to retain such desires. Moreover, if we did not desire to be free from these desires born of addiction, then the problem is reproducible for value-fulfilment theories. The lack of a second-order desire to be free from

the desire born of addiction suggests a level of identification with it. For these reasons, we ought not to be too troubled by alleged counterexamples based on addiction.

Turning to compulsions, desire theories of well-being appear to suggest that fulfilling our compulsions benefits us. This seems counterintuitive regarding compulsions that run counter to our values and interests. Compulsions can also be profoundly strange phenomena. Take, for instance, the call of the void. This is the compulsion that some people experience to jump from a great height when experiencing vertigo (T. Schroeder 2004, 144). It is counterintuitive for a theory of well-being to entail that acting upon this sort of compulsion non-instrumentally improves our well-being. Nevertheless, I do not think that the existence of compulsions fatally undermines desire theories of well-being. The all-things-considered harm explanation invoked to account for real-world cases of addiction can also be applied to compulsions (Hubin 1996, 46). On this view, the benefit of fulfilling the call of the void compulsion is outweighed by the countervailing harms that jumping from large heights typically gives rise to.

Moreover, it may be that we are sometimes mistaken when we conceptualise compulsions as desires. Instead, they may be alternative mental phenomena. For instance, it may be that some compulsions are emotional responses to beliefs, rather than being desires themselves. Consider again the vertiginous compulsion to jump from a great height. In this case, the phenomenology of the compulsion is similar to that of a desire. However, the fact that it typically does not motivate us suggests that we are not dealing with an actual case of desire. Of course, this seems less likely of compulsions that do motivate. Nevertheless, many of the things that are experienced as compulsions may end up not being desires. Donald C. Hubin suggests that, just as we can entertain the propositional content of a belief without endorsing it, it may be that the propositional content of a desire can also be entertained without being endorsed (2003, 325). Just as we are reticent to call the former a belief, we ought not to call the latter a desire. Hubin tentatively claims that this is what is going on in some cases of whims. A similar story could be told of some compulsions. This seems like a plausible claim given the structural similarities between desire and belief (Gregory 2012; Sumner 1996, 124). Consequently, the existence of compulsions does not give us reason to prefer a value-fulfilment theory over a desire theory of well-being.

Finally, and perhaps most fatally, value-fulfilment theories are themselves vulnerable to a version of the problem of unstable desires. As with desire theories of well-being, value-fulfilment theories tend not to specify at which time the intensity of a value determines the extent of its effects on well-being. If these theories take a similar approach to simple

concurrentism and claim that it is the intensity of the value at the time of its fulfilment or frustration that determines the extent of these effects, then a version of the problem of unstable desires is reproducible for these theories. This is because values can also fluctuate in intensity. Moreover, there seems to be nothing that prevents recently acquired pro-attitudes from having the features necessary to qualify as strongly held values. Consequently, this view implies that you benefit more from the fulfilment of recently acquired values than is intuitive; that you benefit less from the fulfilment of long-standing values than is intuitive; and, more generally, that the degree to which you benefit from the fulfilment of a value varies, implausibly, with fluctuations in its strength. Therefore, value-fulfilment theories are vulnerable to a similar critique to that which motivated the jettisoning of simple concurrentism. Of course, it is possible for these theories to avoid this problem by adopting versions of the Averaging and Longevity principles. However, doing so further complicates an already conceptually overburdened family of views. Consequently, we are better off adopting a stability-adjusted desire theory of well-being instead.

Value-fulfilment theories can account for unstable desires and address concerns about alienation. Nevertheless, we have seen that there are good reasons to find these views less attractive than stability-adjusted versions of the desire theory. This is because these views inherit a larger explanatory burden, are not well-motivated, and are susceptible to a version of the problem of unstable desires. For these reasons, I have argued that the appeal of value-fulfilment theories does not eclipse that of desire theories of well-being.

4.5 Prudential reasons and mental disorder

Stability-adjusted versions of the desire theory of well-being have implications for our understanding of prudential reasons. These theories entail that, all things being equal, long-standing desires matter more to well-being than fleeting desires. It follows from this that, all things being equal, it is more instrumentally valuable to preserve rather than lose long-standing desires for attainable outcomes. Moreover, all things being equal, it is more prudent to expeditiously lose desires that have little or no chance of being fulfilled. This is because the effects on well-being of fulfilling or frustrating a desire are amplified by the length of time that the desire is held. Stability-adjusted versions of the desire theory of well-being entail that, to the extent that we are able to engineer our desires, we have prudential reasons to quickly dispense of unachievable desires and to sustain long-standing desires for achievable outcomes. These theories may also give us additional prudential reasons to

develop our capacity to engineer our desires.[57] If we find this type of theory convincing, then we ought to integrate these considerations into our decision-making.[58]

Stability-adjusted desire theories of well-being also have implications for our understanding of the relationship between mental disorder and well-being. Some mental disorders partially disrupt the capacity to maintain long-standing desires. The moral psychology of mental disorder is an underdeveloped field. In Chapter Three I considered how analysing the moral psychology of severe depression can expand our understanding of its effects on well-being. A similar treatment can be extended to other types of mental disorder. For instance, it seems plausible that some mental disorders impede the retention of long-standing desires and facilitate the incubation of fleeting desires. This may be true of borderline personality disorder. A recent literature review of this condition describes it as having the following symptomology:

> 'Sudden shifts in identity, interpersonal relationships, and affect, as well as by impulsive behavior, periodic intense anger, feelings of emptiness, suicidal behavior, self-mutilation, transient, stress-related paranoid ideation, and severe dissociative symptoms' (Leichsenring et al. 2023).

Clearly, this is a somewhat disparate cluster of symptoms. Nevertheless, we can infer from this symptomology that one effect of borderline personality disorder is the partial impairment of the capacity to maintain long-standing desires and goals. This seems to be a consequence of the sudden shifts in identity and impulsive behaviour that characterise this condition. These symptoms combine to partially obstruct the ability of subjects to sustain long-standing desires. Because long-standing desires matter more to well-being than fleeting desires, borderline personality disorder harms subjects by partially diminishing their capacity to retain and fulfil such desires. This way of understanding the harm of borderline personality disorder is not unique to stability-adjusted desire theories of well-being. It is an explanation available to other versions of the desire theory of well-being, and to other

[57] In §3.5 I argued that we do not have prudential reasons to acquire new desires simply because fulfilling them would increase our well-being (Dorsey 2019). However, this does not entail that we have no prudential reason to dispense of unachievable desires, or that we have no prudential reason to preserve those achievable desires that we already have. Consequently, my argument here is compatible with my earlier claim.

[58] Other versions of the desire theory of well-being may entail a similar conclusion based on the opportunity cost of retaining unachievable desires. Nevertheless, stability-adjusted versions of this view can point to the importance of desire longevity as a reason to give this conclusion additional weight in decision-making.

theories that take desire fulfilment and frustration to be a constituent part of well-being. After all, even if the longevity of desires is irrelevant to well-being, it seems likely that the symptomology of borderline personality disorder often impedes the fulfilment of a great many desires. Nevertheless, stability-adjusted desire theories of well-being can point to the partial impediment of the capacity to retain long-standing desires as a unique way in which borderline personality disorder lessens the well-being of the subjects who experience it.[59]

Of course, none of this is to say that the harm of borderline personality disorder is primarily a result of this effect. Indeed, it strikes me that the extreme stigmatisation that often accompanies this disorder plays a larger role in diminishing well-being. As do the inadequate social provisions which often exacerbate symptomology and obstruct recovery. Moreover, the fact that people with borderline personality disorder often experience coexisting mental disorders means that it is difficult to disentangle how exactly this condition affects well-being independently of these comorbidities (Leichsenring et al. 2023). Nevertheless, there is something to be said for the idea that the psychological structure of this condition diminishes the well-being of the subjects who experience it. Stability-adjusted desire theories of well-being can explain one way in which this psychology diminishes well-being.[60]

I have used borderline personality disorder as an example of how desire theories of well-being in general, and stability-adjusted versions of the view in particular, can explain the relationship between the moral psychology of certain mental disorders and the effects that these disorders have on the well-being of the subjects who experience them. There are other mental disorders that may lend themselves to a similar treatment. However, I set aside such considerations here.

[59] There are additional ways in which desire theories of well-being can recognise that the symptomology of borderline personality disorder lessens well-being. For instance, the periodic intense anger that people with this condition tend to experience is likely to lead to some desires over-motivating. This may lead to the fulfilment of those desires at the expense of other stronger desires. Intense emotions often lead to desires having disproportionally strong effects on motivation (§2.4.1).

[60] Value-fulfilment theories can give a richer account of this relationship. They can point to the sudden shifts in identity that are characteristic of borderline personality disorder as leading to a disruption of the valuing attitude. These theories typically define the valuing attitude as a type of pro-attitude that we identify with. For those who have a more amorphous sense of identity it may be difficult to sustain the pro-attitudes that value-fulfilment theories take to be central to well-being.

4.6 Chapter summary

The minimally plausible desire theory of well-being underspecifies the time at which the strength of a desire determines the extent of the effects of its fulfilment or frustration on well-being. According to one version of the theory, simple concurrentism, it is the strength of the desire at the time of its fulfilment or frustration that determines the extent of these effects. However, this theory is undermined by the problem of unstable desires. This problem claims that viable theories of well-being must value the effects of fleeting desires, long-standing desires, and fluctuations in desire strength on well-being in an intuitive way. This is something that simple concurrentism fails to do. For this reason, I argued that more plausible versions of the desire theory of well-being should adopt Averaging and Longevity principles. The resultant stability-adjusted versions of the desire theory of well-being are able to avoid the problem of unstable desires. I then defended this view from two objections and argued that value-fulfilment theories are an inferior way of navigating this problem. Finally, I considered how the adoption of a stability-adjusted desire theory of well-being affects our understanding of prudential reasons and certain mental disorders.

Chapter 5 The Problem of Dead Sea Apples

5.0 Introduction

This chapter considers an objection to desire theories of well-being based upon a certain species of desire fulfilments. Dead Sea apples are desire fulfilments that leave us disappointed and bereft of feelings of satisfaction. It is implausible to claim that Dead Sea apples improve well-being. Yet, this is what desire theories of well-being seem to entail. If this is right, then we have good reason to reject these theories. I argue that purported examples of Dead Sea apples do not describe actual desire fulfilments. Rather, they describe cases where we have a false belief that our desire has been fulfilled. This happens when we have a conjunctive desire that contains some true and some false propositions. For a conjunctive desire to be fulfilled, all of its conjuncts must be true. The fact that purported examples of Dead Sea apples often have some true propositions in their content gives us the false impression that our desire has been fulfilled. Consequently, desire theories of well-being do not entail that Dead Sea apples improve well-being. If this is right, then this argument solves an important problem for desire theories of well-being and enriches our understanding of the moral psychology of disappointment.

§5.1 outlines the problem of Dead Sea apples for desire theories of well-being. §5.2 surveys existent approaches to this problem within the literature and argues that these approaches are inadequate. §5.3 argues that reflections on the structure of desire allow us to explain Dead Sea apples as mere simulacra of desire fulfilments. §5.4 considers how similar observations can explain the effects of pleasant surprises on well-being. §5.5 concludes with a summary.

5.1 The problem of Dead Sea apples

Calotropis procera is a type of flowering plant with fruits so rancid that they have inspired poets, musicians, and philosophers. There is a stark incongruity between the appealing aesthetics and the noxious taste of these fruits. They are sometimes called Dead Sea apples, or Apples of Sodom. John Milton alludes to them when narrating the fall of humanity in his poem *Paradise Lost* (1667):

‘[…] greedily they pluck'd
The Frutage fair to sight, like that which grew

> Neer that bituminous Lake where *Sodom* flam'd;
> This more delusive, not the touch, but taste
> Deceav'd; they fondly thinking to allay
> Thir appetite with gust, instead of Fruit
> Chewd bitter Ashes, which th' offended taste
> With spattering noise rejected: oft they assayd,
> Hunger and thirst constraining, drugd as oft,
> With hatefullest disrelish writh'd thir jaws' (Book 10, lines 560–568).

The clash between the alluring appearance and foul taste of Dead Sea apples makes them a ripe candidate for philosophical analogy. Henry Sidgwick critiques desire theories of well-being by pointing out that desire fulfilments sometimes resemble Dead Sea apples. He writes:

> 'It would still seem that what is desired at any time is, as such, merely apparent Good, which may not be found good when fruition comes, or at any rate not so good as it appeared. It may turn out a 'Dead Sea apple,' mere dust and ashes in the eating' (Sidgwick 1907, 110).

Here Sidgwick draws attention to the intuition that sometimes fulfilling our desires does not improve our well-being because the moment that they are fulfilled we find ourselves disappointed and bereft of positive affective states. We should interpret this passage as describing fulfilled desires that seem to be completely devoid of any benefit to us. Upon fulfilment of these desires, we are left with the impression that our well-being has not been improved. Viable theories of well-being should capture the intuition that desire fulfilments of this sort do not improve well-being. I will refer to desire fulfilments of this sort as Dead Sea apples throughout the remainder of this discussion.[61]

For something to be a Dead Sea apple it must involve both disappointment and the absence of positive affective states. Desire fulfilments that give rise to positive affective states do not

[61] William Lauinger defines Dead Sea apples as simply desire fulfilments that leave us disappointed (2011, 325). However, this definition is too broad. After all, we can be disappointed to receive smaller than expected amounts of something good. Yet, such cases do not pose a challenge to desire theories of well-being. Chris Heathwood defines Dead Sea apples as desires that are 'no longer wanted once they are gotten' (2005, 493). However, this definition similarly underspecifies these phenomena. For instance, we sometimes have second-order desires to no longer have first-order desires for things that we already have. Such cases are unintentionally captured under the umbrella of Dead Sea apples by Heathwood's definition. Conversely, my definition more precisely identifies cases which are of no discernible benefit at all to the agent who experiences them. To my mind, this definition better captures the spirit of Sidgwick's description of the phenomenon.

threaten desire theories of well-being in the same way that Dead Sea apples do. This is because they fulfil standing desires for positive affective states. It seems to be a feature of human psychology that we have such standing desires (§1.3). Consequently, desire fulfilments that give rise to positive affective states provide at least some benefit. However, it is not possible to appeal to this type of explanation for cases of Dead Sea apples. This is because these cases seem to produce no benefit whatsoever.

Not all desires that fail to produce positive affective states qualify as Dead Sea apples. After all, consummatory anhedonia can prevent these mental states from emerging (§3.3), and there are cases where we lack information about whether our desire has been fulfilled or frustrated that similarly produce no such feelings (§1.2). Yet, it is plausible that the fulfilment of such desires does benefit us. One of the virtues of desire theories of well-being is that they do not claim that well-being is entirely dependent upon mental states. Consequently, for something to be a Dead Sea apple it must also give rise to a sense of disappointment. The evaluative judgments of subjects who experience disappointment from desire fulfilments of this sort suggest that their own well-being has not been advanced. When both disappointment and an absence of affective states are apparent, there is a strong intuitive case that the desire fulfilment fails to improve well-being. Desire theories of well-being should be able to capture this intuition.

There are many different types of desires that are sometimes said to be irrelevant to well-being (Bruckner 2016; Heathwood 2005). Nevertheless, what makes Dead Sea apples distinct is that it is the evaluative judgements of the subjects who experience them that suggest that their fulfilment does not improve well-being. Consequently, the case for Dead Sea apples failing to improve well-being derives its force from intuitions internal to the subject who experiences them. This distinguishes the problem of Dead Sea apples from objections based on other types of desire that are sometimes said to be irrelevant to well-being. The case for excluding other types of desires from affecting well-being often involves appeal to evaluative judgements that are external to the subject who experiences those desires (Bruckner 2016, 12). Consequently, they warrant separate consideration.

The existence of Dead Sea apples is a problem for desire theories of well-being. They are alleged counterexamples to the thesis that fulfilling desires always improves well-being. Consequently, if these theories are to remain viable, then they need to have something to say about this problem. The problem of Dead Sea apples can be summarised as follows:

> **P1:** Viable theories of well-being must capture the intuition that Dead Sea apples do not improve well-being.

P2: Desire theories of well-being fail to capture the intuition that Dead Sea apples do not improve well-being.

C: Therefore, desire theories of well-being are not viable theories of well-being.

The next section considers existent approaches within the literature that address how desire theories of well-being can respond to the problem of Dead Sea apples.

5.2 Proposed solutions to the problem of Dead Sea apples

To my knowledge, there are four approaches within the literature to the problem of Dead Sea apples. The first involves appealing to versions of the desire theory of well-being that accept concurrentism and claiming that Dead Sea apples are never produced by desires that temporally overlap with their objects. The second involves appealing to an idealisation version of the desire theory of well-being and claiming that the desires that are relevant to well-being do not produce Dead Sea apples. The third involves restricting which desires affect well-being to only intrinsic desires and claiming that Dead Sea apples only emerge from the fulfilment of instrumental desires. The fourth involves appealing to the view that desires are more fine-grained than is typically assumed and consequently do not produce Dead Sea apples. I argue that all four approaches fail to solve the problem of Dead Sea apples. I then introduce my own proposed solution to the problem in §5.3.

5.2.1 Concurrentism reconsidered

Some versions of the desire theory of well-being accept concurrentism (§4.1). This view claims that only those desires that exist at the same time as their objects non-instrumentally improve well-being (Heathwood 2005, 490). I pointed out earlier that one of the supposed benefits of concurrentism is that it can put forward a proposed solution to the problem of Dead Sea apples. Chris Heathwood argues that this problem does not apply to these views. He writes that:

> ‘Genuine desire satisfaction is had only when the desire remains once its object is gotten. The concurrence requirement ensures that the getting of Dead Sea apples doesn’t improve welfare, since the very reason the thing is a Dead Sea apple is that the desire for it has vanished’ (Heathwood 2005, 493).

According to concurrentism, if a desire does not temporally overlap with its object, then it does not non-instrumentally improve well-being.[62] This may be what is happening in cases of Dead Sea apples. According to this argument, these desires are lost before their objects occur. This gives us the impression that our desire has been genuinely satisfied. However, this impression is mistaken. The fact that desires typically produce feelings of satisfaction when we perceive their fulfilment may lend credibility to the idea that purported cases of Dead Sea apples fail to describe genuine desire satisfaction. On this view, the lack of such feelings is taken as evidence of this claim.

In §4.3 I remained neutral about whether we ought to accept concurrentism. However, if doing so is the only way for desire theories of well-being to solve the problem of Dead Sea apples, then this gives us good reason to prefer this view to its rivals. Nevertheless, I do not think that we should accept concurrentism in response to this problem. There are at least three reasons for this. Firstly, accepting concurrentism means that desire theories of well-being fail to capture important intuitions about well-being. Secondly, there are more attractive alternatives to concurrentism. Thirdly, concurrentism fails to solve the problem of Dead Sea apples. I will consider these arguments sequentially.

Some writers have argued that concurrentism fails to capture important intuitions about well-being. For instance, the fulfilment and frustration of some desires seems to matter to well-being irrespective of whether those desires are lost prior to the time of their objects (Dorsey 2013, 158). This claim is most convincingly made of desires to act in accordance with deeply held values and desires to achieve serious life goals. For instance, a young environmentalist may have a strong desire to abstain from air travel throughout their life. It seems that there is at least some intuitive force behind the idea that the fulfilment or frustration of this type of desire matters to their well-being even if the desire is lost prior to the end of their life. The same seems to be true of serious life goals. For instance, suppose that you lose your long-standing desire to scale Mt. Ijen mere moments before its fulfilment. There seems to be at least some intuitive force behind the idea that completing this goal would nonetheless

[62] Proponents of concurrentism can reach this conclusion in two ways. They can claim that part of the definition of desire fulfilment is that a desire must temporally overlap with its object. Alternatively, they can claim that only those desires that do temporally overlap with their objects are relevant to well-being. The first claim entails the second, but the second claim does not entail the first. If proponents of concurrentism take the second approach, then this makes their view a version of the restricted desire theory of well-being. In this discussion I remain neutral on which of these two approaches are more plausible. Nevertheless, I adopt the language of the second approach by describing the fulfilment and frustration of lost or past desires.

improve how well your life goes for you.[63] Yet, if we accept concurrentism, then we cannot capture this intuition.

Furthermore, accepting concurrentism rules out the possibility of posthumous benefits and harms. It is common for people to have desires that specify certain states of affairs to happen after they die. There is something to the intuition that the fulfilment or frustration of such desires can be relevant to a subject's well-being. Desire theories of well-being that reject concurrentism can claim that desires which do not temporally overlap with their objects can nevertheless affect well-being. This makes posthumous benefits and harms possible (Pitcher 1984). However, if we accept concurrentism, then such desires cannot affect well-being. Of course, not everyone shares the intuition that posthumous events are relevant to well-being. Nevertheless, it strikes me that there is at least some intuitive force behind the idea that desires for things to happen after death are sometimes relevant to well-being. Consider again Sheila in §2.1 who was willing to sacrifice herself in order to further the reconstruction efforts of a war-ravaged city (Darwall 2002, 43–44). Given that such desires can be central life projects, it seems counterintuitive to completely exclude them from ever affecting well-being.

Furthermore, Dale Dorsey observes that it is counterintuitive for desire theories of well-being to treat temporal distance differently from the way in which they treat spatial difference (2013, 158). According to desire theories of well-being, awareness is not a precondition for the fulfilment or frustration of a desire to affect well-being (§1.2).[64] This is what allows these theories to explain why some cases of misperception and deception do not improve well-being. Consequently, these theories allow for things that do not enter into our experience to nevertheless affect our well-being. Given this, it seems strange for these theories not to extend a similar treatment to temporal distance. For instance, suppose that you have a desire that there is life on Mars today and, unbeknownst to you, it turns out that

[63] Stability-adjusted desire theories of well-being entail that we are instrumentally harmed by losing a desire to achieve a serious life goal mere moments before fulfilling it (§4.3). This is because these views claim that the length of time that a desire is held magnifies the extent of its effects on well-being. Consequently, we have strong instrumental reasons not to lose those serious life goals that are achievable. Nevertheless, I think that there is something to the intuition that fulfilling lost serious life goals non-instrumentally improves well-being. Rejecting concurrentism allows us to capture this intuition.

[64] A possible counterexample to this is Chris Heathwood's Subjective Desire Satisfactionism. This view claims that well-being is improved by the subjective perception that a desire has been fulfilled and decreased by the subjective perception that a desire has been frustrated (Heathwood 2006, 559). I argued in §1.2 that this position is better understood as a mental state theory of well-being, rather than a desire theory of well-being.

your desire is fulfilled. According to desire theories of well-being, this means that your well-being is improved despite you not finding out about your desire's fulfilment. Conversely, suppose that you have a desire that there is life on Mars in one hundred years and, unbeknownst to you, it turns out that your desire is fulfilled. If we accept concurrentism, then the fulfilment of this desire does not improve your well-being. It strikes me that desire theories of well-being should treat the effects of both desires similarly. After all, in both cases you do not gain information about the fulfilment of your desire. Nevertheless, according to concurrentism, only the fulfilment of the first desire improves your well-being. This seems wrong. To my mind, these reasons militate in favour of rejecting concurrentism. Nevertheless, in the absence of compelling alternatives, concurrentism may remain the best option available. I turn now to consider the plausibility of versions of the desire theory of well-being that reject concurrentism.

Proponents of concurrentism can point out that, while some desires appear to remain relevant to well-being after they have been lost, clearly this is not true of all past desires. The fact that as a child I wanted to be a zookeeper does not mean that I would be benefited by pursuing that profession now. If rejecting concurrentism means that all such desires are relevant to well-being, then this reduces the theory to absurdity. Consequently, unless we can provide a principled way of distinguishing between which desires remain relevant to well-being after they have been lost, and which desires are no longer relevant, then we are better off simply accepting concurrentism.

One approach claims that only those desires which are not conditional on their own persistence continue to be relevant to well-being after they have been lost (Parfit 1984, 151). A desire is conditional on its own persistence when it specifies the continuation of the desire as part of its content. Desires of this type require themselves to exist at the time of their object in order to be fulfilled. Conversely, desires are unconditional on their own persistence when they do not specify the continuation of themselves as part of their content. For instance, people often write a last will and testament to communicate how they would like their assets to be divided after their death. Presumably, the desire that motivates this action is one that is not contingent upon its own persistence. After all, the subject knows that they will be dead at the time of the object of their desire. By distinguishing between desires that are conditional on their own persistence and those that are not, desire theories of well-being can put forward a plausible explanation about which desires remain relevant to well-being after they have been lost and why.

Versions of the desire theory of well-being that reject concurrentism are sometimes criticised for struggling to provide an account of the temporal location of benefits and harms (Baber 2010, 257). Failure to provide such an account would be a disaster for the resultant theory (Dorsey 2013, 154). Fortunately, there are several candidates for this location. The time-of-object view claims that well-being is affected at the time that a desired state of affairs occurs or fails to occur (Dorsey 2013, 156). This can be after the desire has been lost. Conversely, the time-of-desire view claims that well-being is affected at the time that the desire is held (Dorsey 2013, 158–159). This can be prior to the desired state of affairs occurring or failing to occur. In §1.2 I pointed out that the time-of-desire view better captures the resonance requirement. For this reason, this strikes me as the most plausible position for views that reject concurrentism to take. Regardless of which position we take, concurrentism does not have a monopoly on the conceptual space of temporal locations for benefits and harms. Therefore, versions of the desire theory of well-being that reject concurrentism should not be dismissed for this reason.[65]

Finally, it is worth noting that some writers are hesitant to accept the idea that past desires matter to well-being because they assume that doing so means that past desires matter to well-being as much as present desires (Parfit 1984, 152–153). However, it is perfectly possible to claim that a subject's past desires matter less to their well-being than their present desires (Vorobej 1998, 314–315). Moreover, as we have already seen, prudential reasons are not tied to well-being maximisation in all contexts (§3.5). For instance, we do not have prudential reasons to acquire new desires simply to fulfil them, or to engineer our desires so that they match the world as it is (Dorsey 2019). This is true even though doing so would improve our well-being. Consequently, it may be that the prudential reasons generated by past desires do not carry the same normative weight as those generated by present desires. Considerations of this sort may ameliorate some of the concerns that have motivated some writers to reject alternatives to concurrentism.

Moreover, even if we are unmoved by arguments of this sort, there is reason to think that an appeal to concurrentism does not in fact solve the problem of Dead Sea apples. The idea that Dead Sea apples happen only when our desire vanishes prior to its fulfilment is a neat solution to the problem. However, it is not a convincing explanation of all cases of Dead Sea apples. Sometimes we appear to have fulfilled desires that persist despite giving rise to no

[65] Eden Lin puts forward an alternative theory whereby sometimes well-being is affected at the time of desire, and at other times at the time of object (2017b, 179–180). I will not consider this view here.

feelings of satisfaction and which continue to leave us disappointed (Lauinger 2011, 331–332). This is a distressing and disorienting experience. Yet, it does not seem to be particularly uncommon. We frequently continue to want the thing that seems to be leaving us disappointed and unsatisfied. Lauinger calls this 'the overlap problem' for the concurrentism response. This problem renders this response otiose.

The idea that the problem of Dead Sea apples can be traversed through appeal to concurrentism may initially appear to be appealing. Nevertheless, there are three reasons to think that this approach fails. Firstly, concurrentism fails to capture important intuitions about well-being. Secondly, there are more attractive alternatives to concurrentism. Thirdly, concurrentism fails to solve the problem of Dead Sea apples. Therefore, we will need to look elsewhere for a solution to this problem.

5.2.2 Idealisation theories

According to some versions of the desire theory of well-being, it is the fulfilment and frustration of a subject's counterfactual desires that determine their well-being, rather than that of their actual desires. These theories differ in how they arrive at which counterfactual desires are relevant to well-being. Views of this sort are called idealisation theories. This is because these theories claim that the desires which are relevant to well-being are those that we would have, or that we would want ourselves to have, in an idealised set of circumstances. It may be that these counterfactual desires do not produce Dead Sea apples. If this is correct, then the problem of Dead Sea apples does not undermine versions of the desire theory of well-being that accept an idealisation account.

Henry Sidgwick discusses a view of this sort in response to the problem of Dead Sea apples. He considers the view that the good for a person is 'what would be desired, with strength proportioned to the degree of desirability, if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition' (1907, 111). Richard Brandt advances a similar position. He argues that only those desires that would survive a vivid appreciation of their consequences and which are informed by all the appropriate available information are relevant to well-being (Brandt 1979, 110–115). Peter Railton similarly argues that 'an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his

circumstances, and entirely free of cognitive error or lapses of instrumental rationality' (1986, 16).

Regardless of the specifics of whichever idealisation theory we favour, there are several reasons to think that this sort of approach is not a viable solution the problem of Dead Sea apples.[66] David Enoch argues that idealisation theories are an intolerably *ad hoc* way of achieving extensional adequacy (2005, 760–761). He points out that these theories specify counterfactual desires that are very different from our everyday desires. Consequently, he thinks that there is no principled reason to accept views of this sort beyond their ability to achieve extensional adequacy. This is not a sufficient reason to accept a theory. After all, extensional adequacy can be achieved through a messy proliferation of arbitrary conjuncts and disjuncts added to any theory in order to patch up its counterintuitive implications (Enoch 2005, 766–769). He argues that unprincipled approaches of this sort are deeply implausible even if they do manage to achieve extensional adequacy. This is because the method by which extensional adequacy is achieved must also be independently plausible (§1.1).

Enoch's conclusion strikes me as somewhat overstated. Idealisation theories are principled when they start from the observation that desires are typically a good basis for understanding well-being. From this premise proponents of these views can observe that, under some specific circumstances, we form desires that are bad for us and fail to form desires that are good for us. Consequently, there is at least something principled about the thought that only the desires that we would have outside of those circumstances are what determines well-being. Nevertheless, idealisation theories are less principled when they specify very different counterfactual circumstances from our actual circumstances. In cases like this, it seems that the premise that desires are a good basis for understanding well-being is abandoned in favour of an external standard of well-being.

Part of the appeal of desire theories of well-being is their ability to capture the intuition that our well-being is something that must resonate with our own subjective attitudes (§1.2). If there are devastating counterexamples to this thesis, then this seems to be a reason to reject the underlying approach, rather than to amend it through appeal to idealisation. Indeed, there seems to be some sleight of hand at play in the idealisation theories of Sidgwick, Brandt,

[66] Given the sheer breadth of literature on this topic, I limit myself to engaging selectively with those parts of it that are most pertinent to the problem under consideration. David Enoch has a detailed discussion of further reasons why we ought to reject idealisation theories (2005). David Sobel contests some of these arguments in a response to Enoch (Sobel 2009).

and Railton. The counterfactual desires produced by these idealisation processes are likely to be so different from our actual desires that the resultant theory surely fails to capture any plausible interpretation of the resonance requirement (Raibley 2010, 594). Consequently, these theories seem to share more in common with those objective list theories that postulate a range of alienated well-being goods than they do with standard versions of the desire theory of well-being (§1.4).

This does not establish that all forms of idealisation are unviable. Moderate types may not be. For instance, approaches that idealise certain pro-attitudes in light of more fundamental pro-attitudes do not seem to be intolerably *ad hoc*, and nor do they alienate an individual from their own good (Dorsey 2021, 143; Dorsey 2017; Noggle 1999, 321–322). However, moderate idealisation processes offer no reprieve from the problem of Dead Sea apples. This is because Dead Sea apples do not always arise from ill-informed or ill thought through desires. Indeed, it seems likely that even perfectly rational and adequately informed people are susceptible to their desires leading to feelings of disappointment and dissatisfaction (Baber 2010, 251).

Moreover, even if this were not the case, the process by which idealisation theories identify which counterfactual desires are relevant to well-being is beset with procedural difficulties. This is especially true of the views put forward by Sidgwick, Brandt, and Railton. For instance, the order of presentation of facts typically plays a role in shaping the desires that people acquire when they are given more information (Loeb 1995, 12). This facet of human psychology is a problem for idealisation theories. It shows that full information will produce different results depending upon how that information is presented. This raises the question of how idealisation theories should specify the presentation of information in the idealisation process. To my mind, there is no non-arbitrary way of doing so. Railton notes this problem but argues that this effect is likely to be minimal in cases where large amounts of information is provided (1986, 21). This brings us to a related problem.

Idealised versions of ourselves presumably share our cognitive limitations. If this is right, then they share our inability to process full information and maximal vividness in complex situations (Loeb 1995). Therefore, even if Railton is right that more information will minimise the effects that the order of presentation of information have on desires, this is no good if we lack the cognitive architecture to process that information. Consequently, there is an additional reason to think that the idealisation processes favoured by Sidgwick, Brandt, and Railton cannot solve the problem of Dead Sea apples. Of course, proponents of idealisation theories can respond to this concern by specifying perfect cognitive capacities

on the part of the idealised versions of ourselves. However, this approach further alienates the concerns of our actual selves from those of our idealised counterparts. Once idealisation theories interfere with our cognitive and psychological architecture, it is unclear how the resultant counterfactual desires remain our own in any meaningful way.

Finally, it is worth noting that even if these difficulties can be traversed, and idealisation can solve the problem of Dead Sea apples, there are still independent reasons to reject this approach. Idealisation is often appealed to in order to connect individuals more clearly to the things that are good for them. Yet, these theories sometimes have the opposite effect. They sometimes sever the connection between an individual and their good (Keller 2009, 658). For instance, sometimes people have desires to engage recklessly in high-risk behaviours involving drugs and alcohol. Many of these desires would not survive full acquaintance with the facts about the negative effects on health, finances, and life prospects that such behaviours typically engender. Nevertheless, it is absurd to claim that fulfilling those desires is of no benefit whatsoever to the subject that has them. Of course, these actions may turn out to be all-things-considered bad for them over time. But popular idealisation theories seem to entail that their fulfilment does not improve well-being at all. If we are tempted to adopt an idealisation theory in order to remedy the problem of Dead Sea apples, then this iatrogenic harm should dissuade us from doing so. The well-being of imperfect beings should reflect the imperfections that constitute those beings.

Idealisation theories have been influential in the literature on desire theories of well-being. Nevertheless, there are several reasons to reject them as solutions to the problem of Dead Sea apples. Firstly, some of these theories are intolerably *ad hoc*. Secondly, many of these theories forfeit the ability to capture plausible interpretations of the resonance requirement. Thirdly, there are procedural difficulties that make many of these theories implausible. Fourthly, many of these views counterintuitively exclude some of our actual desires from affecting well-being. Consequently, we should set aside the idea that the desires of idealised versions of ourselves are what determines our well-being. Instead, we should look elsewhere for solutions to the problem of Dead Sea apples.

5.2.3 Only intrinsic desires count

Donald Hubin has developed an account of the nature of reasons that contains insights relevant to the problem of Dead Sea apples. He argues that our reasons for action are determined by our intrinsic motivations. He defends this view from the objection that some

of our actual motivations are too defective a basis for our reasons (Hubin 1996, 40). He summarises this objection as follows:

> 'The view that one has reason to do whatever one is moved to do is too crude. Sometimes our motivation would not hold up to a moment's reflection: appreciation of the cause of our motivation, its nature, or the real effects of acting in accordance with it may annihilate the motivation' (Hubin 1996, 31).

This argument has parallels with the objection that some of our actual desires are too defective to base our well-being upon. Both arguments leverage cases where our attitudes are ill-informed or ill thought through in order to claim that these attitudes are an implausible basis for our theory of reasons or well-being. Hubin responds to these concerns by claiming that only our intrinsic motivations generate reasons for action (1996, 33). He argues that intrinsic motivations are not susceptible to the same defects as instrumental motivations. On this view, many defective motivations are explained as instrumental motivations based upon false beliefs about what would satisfy intrinsic motivations (Mendola 2009, 149). Others can be explained as ill thought through instrumental motivations that fail to satisfy intrinsic motivations. If this is right, then we can preserve the claim that our reasons for action are based upon our actual motivations (Hubin 1996, 45).

Hubin's argument is premised upon a distinction between intrinsic and instrumental motivations (1996, 44). One way of making this distinction claims that a motivation is instrumental when its existence is completely dependent upon other motivations. Conversely, motivations are intrinsic when their existence is not completely dependent upon other motivations. An advantage of this approach is that it can capture Aristotle's insight that some ends are both instrumentally and intrinsically valuable (1097a30–35). This way of distinguishing between instrumental and intrinsic motivations can claim that motivations generate reasons to the extent that they have intrinsic components. Regardless of our favoured way of distinguishing between instrumental and intrinsic motivations, it seems plausible that such a distinction can be made.[67]

[67] Mark Murphy offers an alternative view. He argues that we do not need to postulate the existence of instrumental desires to explain action (Murphy 1999, 252–256). He claims that what we think of as instrumental desires are simply derivative effects of basic desires. If this is right, then we do not need to restrict which desires count to only intrinsic desires because all desires are intrinsic. Since I do not rely upon the claim that only intrinsic desires matter to well-being in order to solve the problem of Dead Sea apples, I set aside these issues here.

This chapter is not concerned with the correct account of reasons. However, it is possible to borrow elements of Hubin's argument in order to explain how desire theories of well-being can solve the problem of Dead Sea apples. We can do so by claiming that Dead Sea apples never emerge from the fulfilment of intrinsic desires. Rather, they arise when we fulfil an instrumental desire. On this view, disappointment signals that a fulfilled desire was not intrinsic (Sobel 2009, 350). Given that this view does not take the fulfilment of instrumental desires to improve well-being, our resultant theory has no problem explaining why Dead Sea apples do not improve well-being.

It is often taken for granted that desire theories of well-being should exclude instrumental desires from affecting well-being. Most contemporary commentators, proponents, and critics of desire theories of well-being often simply assume this restriction in their definitions of these theories.[68] There are good reasons for this. It has been pointed out that, without this restriction, these theories are committed to a counterintuitive double counting problem (Mendola 2009, 149; Murphy 1999, 253). The unrestricted view entails that the more instrumental desires we fulfil on our way to fulfilling an intrinsic desire, the better it is for our well-being. Given that the existence of instrumental desires is entirely dependent upon the intrinsic desires that they are related to, this seems wrong. For this reason, it strikes me that the balance of reasons suggest that we ought to adopt the view that only intrinsic desires count towards well-being.

However, this restriction fails to solve the problem of Dead Sea apples. Unfortunately, Dead Sea apples do not solely blossom out of the fulfilment of instrumental desires. It is not uncommon to find oneself confronted by disappointment and a lack of feelings of satisfaction when fulfilling even long-standing intrinsic desires. Take, for instance, this well-known description of Alexander the Great in his moment of triumph, 'When Alexander saw the breadth of his domain, he wept, for there were no more worlds to conquer'.[69] I interpret Alexander's tears to be tears of sorrow. They serve as evidence of the intrinsic nature of his desire. And yet, the fulfilment of this desire seems to have made him no better off. We need not be despots or warlords to relate to Alexander. The emotions ascribed to him exemplify something which we are perfectly capable of introspectively detecting. Dead Sea apples can

[68] Heathwood makes some tentative dissenting remarks in one of his papers (2019, 669). Some other writers remain strategically ambiguous as to whether we ought to accept this view.

[69] The quote is sometimes misattributed to Plutarch. However, it appears to actually originate from the antagonist in the movie *Die Hard* (1988), Hans Gruber. I am using it for its stark imagery, rather than for its historical accuracy.

arise from even our intrinsic desires. Consequently, the question of whether we ought to restrict which desires affect well-being to instrumental desires is moot.

It may be thought that proponents of this approach can simply reiterate the claim that Dead Sea apples arise solely from instrumental desires. They can do so by claiming that, despite appearances to the contrary, Dead Sea apples indicate that our desire was not really intrinsic after all. There is a rich history of writers claiming that our introspective endeavours are an unreliable guide to discerning between instrumental and intrinsic desires. There seem to be few, if any, desires that are invulnerable to this line of argument. For instance, Aristotle can be interpreted as claiming that all reasons for action are instrumental in service of the single end of achieving eudaimonia (1094a20–25). John Stuart Mill is sometimes interpreted as claiming that all of our desires are instrumental in the pursuit of the intrinsic desire for pleasure (1861). Given these precedents, proponents of this solution to the problem of Dead Sea apples can claim that we are incorrect when we attribute an intrinsic nature to desires that produce Dead Sea apples.

Nevertheless, this approach requires a dramatic revision of our intuitions about the sorts of things that count as intrinsic desires. We should be willing to revise aspects of our moral psychology in light of the problem of Dead Sea apples, but the idea that these desires are always instrumental is implausible. Picture again the crestfallen Alexander. Ashen and lachrymose at the apex of his success. There is no obvious further desire at which his conquest is in the service of. Postulating further hidden intrinsic desires is an implausibly *ad hoc* manoeuvre solely made to rescue desire theories of well-being from the problem of Dead Sea apples. For this reason, we should reject this approach.

5.2.4 The fine-grained response

William Lauinger considers and rejects a potential solution to the problem of Dead Sea apples that he terms the 'fine-grained response' (2011, 327–329). This approach claims that:

> 'People never desire things like having such-and-such a job or being in such-and-such a romantic relationship. All that people ever desire are certain aspects – indeed, certain positive aspects – of having such-and-such a job and of being in such-and-such a romantic relationship' (Lauinger 2011, 328).

The fine-grained response requires us to revise our everyday notion of desire into a series of individuated desires for specific aspects of states of affairs. For instance, perhaps when we

say that we desire a specific job, we are instead expressing our desires for financial security, the admiration of our peers, and enjoyment. On this view, when our new job turns out to be a Dead Sea apple, it is because we have failed to get those aspects of it that we actually desired. Consequently, Dead Sea apples do not involve desire fulfilments at all. Rather, they involve only the illusion of desire fulfilment. If this is right, then the problem of Dead Sea apples does not undermine desire theories of well-being. This is because Dead Sea apples are no threat to the thesis that fulfilling desires always improves a subject's well-being.

However, Lauinger points out that the fine-grained response requires us to revise significant intuitions about how desires are structured (2011, 328–329). According to the new picture, desires are always fine-grained. Yet, this is a profoundly implausible picture. For instance, it seems overwhelmingly intuitive that sometimes we simply desire a certain job or relationship, rather than only certain aspects of those things. To claim that there are always myriad individuated desires requires us to revise important intuitions about moral psychology. Indeed, this approach does sufficient violence to our intuitions so as to undermine the viability of whichever theory of well-being requires us to take this position. This alone is sufficient reason to reject it.

A slightly more plausible version of this argument claims that our desires for specific states of affairs do exist, however their strength pales in comparison to background desires for certain aspects of those things. On this view, we are not mistaken as to the content of our desires, only the extent of their significance. Consequently, it may be that when the new job turns out to be a Dead Sea apple it is because, while we did get something that we wanted, in doing so we frustrated other more central background desires. This argument entails that Dead Sea apples are all-things-considered harms. However, this approach also fails to solve the problem of Dead Sea apples. This is because the definition of a Dead Sea apple specifies that these phenomena provide no discernible benefit whatsoever. Desire theories of well-being have no problem explaining desire fulfilments that make us worse off on aggregate. What troubles these theories about Dead Sea apples is that these experiences seem to be without any benefit whatsoever to us. Therefore, the fine-grained response also fails to solve the problem of Dead Sea apples.

5.3 The conjunctive desire response

The existent approaches in the literature fail to solve the problem of Dead Sea apples. However, there is a more promising approach. This view claims that we mistake purported examples of Dead Sea apples for desire fulfilments because we have an oversimplified understanding of the content of these desires. On this view, Dead Sea apples are frustrated conjunctive desires that contain some true propositions in their content. A conjunctive desire is one that can be accurately represented as having multiple propositions in its content which are conjoined by an "and" logical operator. For a conjunctive desire to be fulfilled, all of its conjuncts must be true. The fact that these desires contain some true propositions gives us the false impression that our desire has been fulfilled. This argument denies that purported examples of Dead Sea apples describe actual desire fulfilments. Consequently, desire theories of well-being have no problem in explaining why these examples fail to improve well-being. Call this view the 'conjunctive desire response' to the problem of Dead Sea apples.

Conjunctive desires are commonplace. Earlier I argued that one of the harms of severe depression emerges from its suppression of pleasure from emerging (§3.3). This inevitably frustrates conjunctive desires to enjoy certain states of affairs. Desires for enjoyment have two distinct propositions in their content. They specify an object, and they specify that the subject with the desire takes pleasure in the existence of that object. If consummatory anhedonia blocks pleasure from emerging, then conjunctive desires for enjoyment are necessarily frustrated. This is the case irrespective of whether the other proposition specified in their content is true. This is a precedent for my view that purported examples of Dead Sea apples describe frustrated conjunctive desires with some true conjuncts.

This explanation is made clearer when applied to purported examples of Dead Sea apples. William Lauinger provides one such example: 'Dennis wants to work at a certain law firm and then gets the job, only to find that he hates it. Here his desire is fulfilled, but his well-being is not advanced' (2011, 327). We can characterise Dennis' predicament in my terms in the following way: 'Dennis has a conjunctive desire to work at a law firm *and* take pleasure in this work. He gets the job, only to find that he does not take pleasure in it. Here his conjunctive desire is frustrated, and his well-being is not advanced.' This characterisation differs from Lauinger's by redescribing Dennis' desire as conjunctive and containing some true and some false propositions. Nevertheless, this strikes me as an intuitive account of Dennis' psychology. After all, few people desire to get a new job irrespective of whether they dislike that job when they get it. For those who do have a desire of this sort, it seems

unlikely that its fulfilment would give rise to the feelings of disappointment that are characteristic of Dead Sea apples. Consequently, Lauinger's example is well explained by the approach that I have outlined.

The conjunctive desire response to the problem of Dead Sea apples can be fruitfully contrasted with the fine-grained response considered in §5.2.4. Both arguments involve redescribing purported examples of Dead Sea apples in order to paint a picture of frustrated desires that we mistakenly identify as fulfilled. The fine-grained response does so by claiming that our desires are only for specific aspects of situations. Conversely, the conjunctive desire response claims that our desires can be for complex relations. Whereas the former approach fails by implausibly entailing that we never desire things such as jobs or relationships, the latter captures the intuition that we often have such desires. In this way, the conjunctive desire response better explains the complexity of our mental lives.

A possible objection to the conjunctive desire response claims that it is impossible to desire P & Q conjunctively without independently desiring P and independently desiring Q. If this is right, then the conjunctive desire response cannot solve the problem of Dead Sea apples. This is because desire theories of well-being would entail that the fulfilment of the independent conjuncts benefits us. Consequently, the theory would be unable to capture the intuition that purported examples of Dead Sea apples are of no benefit to us. However, it strikes me as relatively common to desire a conjunction without desiring either of its conjuncts separately. For instance, it is possible to desire salt and tequila conjunctively without desiring either conjunct separately. Or, to take another example, it is possible to desire to run and listen to music conjunctively without desiring either conjunct separately. Examples of this sort illustrate that to desire P & Q conjunctively does not entail a desire for either P or Q independently. Consequently, the conjunctive desire response is not undermined by this objection.

According to my argument, we sometimes misapprehend the content of our own desires. One reason to think that this happens concerns what Delia Graff Fara calls 'the problem of underspecification' (2013). She points out that sometimes our verbal reports of the content of our desires fail to specify the full range of conditions under which our desire would be fulfilled. For instance, if I say that 'I desire to eat fruit', then in ordinary cases verbal reports of this sort underspecify the desire's content. This is because it is not the case that any type of fruit in any quantity is capable of fulfilling the desire expressed. For instance, mealy tomatoes would not satiate it. Nor would a single blueberry. One way of responding to this problem is to claim that this verbal report implies that only certain fruits in only certain

quantities are capable of fulfilling the desire (Fara 2013, 255–256). However, this is not a convincing explanation in all cases of verbal underspecification. This is because we sometimes fail to introspectively identify the full conditions that are required to fulfil a desire. Consequently, it seems unlikely that our verbal reports imply conditions that we do not ourselves introspectively detect. The conjunctive desire response can explain what is happening in cases like this. On this view, our desires do specify the range of conditions required for their fulfilment. However, we sometimes have an oversimplified understanding of the content of these desires. Consequently, sometimes our verbal reports reflect our own oversimplified understanding of this content.

It is worth noting that there may be external impediments that obstruct how efficient we are at identifying the content of our own desires. For instance, Chris Heathwood has observed that there seems to be a relationship between advertising and the emergence of Dead Sea apples (2005, 493–494). This seems right. A pernicious effect of advertising seems to be its ability to obscure the content of our desires and thereby motivate us to acquire things that we do not really want. We should strive to understand the moral psychology behind this effect. The argument advanced in this section can be extended to supply an explanation. On this view, an effect of advertising is that it leads us to oversimplify the content of our own desires. Consequently, advertising can beguile us into pursuing outcomes that turn out to be disappointing and unsatisfying. This is because it has led us to attend selectively to only parts of the content of our conjunctive desires. Of course, this is not to say that this is the only, or even most pertinent, way in which advertising immiserates us. Nevertheless, the conjunctive desire response allows us to explain this effect of advertising.

Some readers may be concerned that the argument of this section fails to establish that all cases of Dead Sea apples emerge from frustrated conjunctive desires. It is perfectly possible for a reader to accept that sometimes Dead Sea apples emerge from frustrated conjunctive desires, while claiming that at other times they emerge from fulfilled desires. If this is right, then the problem of Dead Sea apples is not solved. I confess that there is no analytic proof that can be given at this point. All we are left with is inference to the best available explanation. However, there are good reasons to think that the conjunctive desire response provides the most plausible explanation. This is because there are independent reasons to think that Dead Sea apples do not describe actual desire fulfilments. Our desires typically produce feelings of satisfaction when we perceive their fulfilment. Moreover, disappointment is an atypical response to perceiving the fulfilment of our desire. Given that these responses are unusual, when both are present, it seems that we are experiencing a very

atypical case of desire fulfilment. However, these reactions are relatively common responses to the frustration of our desires. This suggests that purported examples of Dead Sea apples may actually describe frustrated desires, rather than desire fulfilments.

My argument is premised upon the idea that human psychology is structured in such a way that disappointment and a lack of feelings of satisfaction do not emerge together in response to perceiving that we have fulfilled our desire. If this is right, then reflections on the nature of desire can solve a serious problem for desire theories of well-being. However, it may be that there remain some residual cases of purported Dead Sea apples that are not well explained by this approach. If this turns out to be true, then the problem of Dead Sea apples is not completely solved by the conjunctive desire response. However, even if this is correct, my argument at least establishes that Dead Sea apples are far rarer than first appears to be the case. If there remain a small number of Dead Sea apples in non-standard cases, then this is not a sufficiently strong reason to reject desire theories of well-being. After all, we should not expect viable theories of well-being to capture our intuitions in every case. It is enough that they do so in the bulk of them.

The problem of Dead Sea apples is a serious challenge to the viability of desire theories of well-being. This objection is unique for deriving its force from intuitions internal to the subject who experiences the desires said to fail to improve well-being. Yet, the findings of this section can solve this problem. On this view, we are mistaken when we think that purported examples of Dead Sea apples describe actual desire fulfilments. Rather, these examples describe frustrated conjunctive desires with some true and some false propositions. The fact that these desires have some true conjuncts gives us the mistaken impression that our desire has been fulfilled. This explanation also enriches our understanding of the moral psychology of disappointment. Sometimes we are left disappointed because we have misapprehended the content of our desire. In such cases, we have mistaken a conjunctive desire for a simpler desire. I turn now to consider how observations made in this section can be applied to explain the relationship between pleasant surprises and well-being.

5.4 The moral psychology of pleasant surprises

Sometimes we are pleasantly surprised to get something that we did not previously want. There are three types of pleasant surprises that are relevant to this discussion. The first happens when we encounter a pleasant surprise that we have no pre-existing desire for. The

second happens when we encounter a pleasant surprise that seems to satisfy a pre-existing desire for something relevantly similar. The third happens when a pleasant surprise emerges from an object that we had a pre-existing aversion towards. Desire theories of well-being should have something to say about how each of these different types of pleasant surprise affect well-being. I consider them sequentially.

The first type of pleasant surprise is the easiest for desire theories of well-being to deal with. Sometimes we are pleasantly surprised by something that we had no pre-existing desire for. For instance, suppose that Clare has never tasted a durian before and has no pre-existing desire to try one. Upon trying durian, Clare enjoys it very much. Desire theories of well-being should capture the intuition that Clare's well-being is improved by this experience. An intuitive way of understanding the moral psychology behind this phenomenon is to claim that pleasant surprises happen when we acquire a new desire for an object at the time that we find ourselves encountering it. This means that Clare is benefitted by the fulfilment of her newly acquired desire for durian. We can supplement this explanation by claiming that often when we find ourselves pleasantly surprised it is because the object fulfils standing desires that we have for positive affective states (§1.3). Consequently, the lack of a specific *ex ante* desire does not mean that desire theories of well-being cannot account for standard cases of pleasant surprises. On this view, well-being is improved by the desire that we acquire for the object at the moment that we encounter it, and by the fulfilment of standing desires for positive affective states.

However, this picture becomes more complicated in situations when we consider the second type of pleasant surprise. Sometimes we find ourselves satisfied with something relevantly similar to the object of our desire. For instance, online shoppers of major supermarkets will be familiar with the item substitutions that frequently blight the deliveries of these goods. However, occasionally the substituted item is just as satisfying as the *ex ante* object of our desire. For instance, if I receive a bottle of malbec instead of the bottle of shiraz that I had a pre-existing desire for, then I seem to be made no worse off for the substitution. An intuitive explanation of what is happening in these cases is that the previous desire is simply lost in favour of a new equally strong desire for the new item. Perhaps these specific desires also exist alongside a broader background desire for wine that can be fulfilled by either object.

While this explanation makes sense, it is one that proponents of desire theories of well-being should be reluctant to accept. This is because, according to the most plausible versions of these theories, the length of time that a desire is held magnifies the extent of the effects of its fulfilment or frustration on well-being (§4.3). Therefore, if an old desire is replaced with

a new desire at the moment that we are pleasantly surprised by an object relevantly similar to our *ex ante* desire, then these theories imply that we are benefitted less by the substitution than we would have been by the originally desired object. This seems wrong. If I desire a bottle of shiraz, but am equally pleased by a bottle of malbec, then intuitively I am equally benefitted by both. Moreover, even setting aside prior axiological commitments, there is something to the intuition that our original desire was fulfilled, rather than lost and replaced by a similar desire. Our moral psychology should capture this intuition.

One way of addressing this concern is to claim that a desire can survive minor changes in its content and still count as the same desire – at least for the purposes of determining its effects on well-being. If this is right, then when I am confronted with malbec rather than shiraz, my desire undergoes a minor change in its content. This means that my *ex ante* desire is fulfilled by the relevantly similar object. Consequently, I am not benefitted less by getting malbec than I would be by getting shiraz. Both objects are capable of fulfilling the same *ex ante* desire. This means that the existence of this species of pleasant surprises does not undermine stability-adjusted versions of the desire theory of well-being. Moreover, it captures the intuition that sometimes desires can be fulfilled by objects relevantly similar to those specified in their *ex ante* content.

This approach raises the question of how we determine what counts as a minor change in content. I will not attempt to provide an exhaustive account of this phenomenon. I limit myself to proposing a way of understanding those minor content changes that constitute pleasant surprises of this sort. Sometimes the content of the desire is modified through addition of a new proposition which is conjoined to the wider content of a desire with an "or" logical operator. The resultant disjunctive desire is fulfilled if either of its disjuncts are true. This seems to be what is happening in cases where a previously undesired object satisfies a pre-existing desire for a relevantly similar object. Such changes should be conceptualised as minor changes in content, rather than new desires – at least for the purposes of determining well-being. This explains how desires can be satisfied by objects that are relevantly similar to their *ex ante* objects. If this is right, then I am equally benefitted by receiving shiraz as I would be by receiving malbec.

Another species of pleasant surprise occurs when we have an *ex ante* aversion to an object that turns out to pleasantly surprise us. Wayne Sumner points out that pleasant surprises of this sort are a problem for desire theories of well-being. He writes of this:

> 'Having never heard bluegrass, I chance on a band playing in the park and find that I like it. Having nursed a long-standing suspicion of the Mediterranean, I am persuaded

> against my better judgement to holiday there and have a wonderful time. In neither case did I have an antecedent desire for the state of affairs which, as it turns out, enhances my well-being. But then, just as satisfying a desire on my part is not logically sufficient for something to benefit me, it is not logically necessary either' (Sumner 1996, 133)

Sumner's experience of bluegrass is explainable as the first type of pleasant surprise discussed. On this view, he acquires a desire for bluegrass upon encountering it, and its lilting tonality fulfils standing desires for positive affective states. However, his description of his Mediterranean holiday is more worrying for proponents of desire theories of well-being. In this case, he has an *ex ante* aversion to something that turns out to benefit him. Proponents of desire theories of well-being should have something to say about such cases.

Fortunately, there are a couple of explanations available. Firstly, it may be that in such cases the frustration of the specific aversion to the Mediterranean holiday is outweighed by the benefits of the other desires that such a holiday likely fulfils. On this view, experiences of this type are all-things-considered benefits to the person who has them. Nevertheless, this explanation cannot be applied to all examples of this type of pleasant surprise. Sometimes the object that we have an *ex ante* aversion towards seems to be of no harm whatsoever to us. In this respect, it is the inverse of a Dead Sea apple. Rather than an absence of anticipated benefits, there is instead an absence of anticipated harms.

The parallel structure of Dead Sea apples and this type of pleasant surprise means that we can appeal to a parallel argument to explain these cases. On this view, some aversions are conjunctive aversions. These are aversions that can be accurately represented as containing multiple propositions which are conjoined by an "and" logical operator in their content. When our aversion fails to trigger the expected negative affective and evaluative responses, it is because only parts of its content turn out to be true. For instance, Sumner's aversion to holidaying in the Mediterranean can be redescribed using this approach as an aversion to holiday *and* being miserable on that holiday. This seems plausible. After all, few people have aversions to holidays irrespective of the feelings of dissatisfaction, stress, or boredom that they anticipate such an experience will involve. If this is right, then desire theories of well-being can appeal to a mirror of the argument made in §5.4 to explain how this type of pleasant surprise affects well-being. This explains why we are not harmed by such pleasant surprises. Moreover, we can explain the benefit of such experiences as emerging from the newly acquired desires that are acquired upon encountering a pleasant surprise, and by the fulfilment of standing desires for positive affective states. This explanation is the same that

was appealed to in order to explain the first type of pleasant surprise discussed in this section. It works just as well here.

Desire theories of well-being may be thought to be unable to account for pleasant surprises. However, the arguments of this section show that this is not the case. Some cases of pleasant surprises can be explained by the fulfilment of desires acquired at the time that we encounter the pleasantly surprising object, and by the fulfilment of standing desires for positive affective states. Other cases of pleasant surprises can be explained through minor changes in the content of pre-existing desires. The resulting disjunctive desires can be fulfilled by the newly specified object in their content being true. Finally, remaining cases of pleasant surprises can be explained as conjunctive aversions that fail to harm us because only part of their content is true. The benefits of these pleasant surprises emerge in the same way as they do in the first type of pleasant surprise considered – through the acquisition of new fulfilled desires, and by the fulfilment of standing desires for positive affective states. The findings of this section lend support to those of §5.3. They show that an appeal to conjunctive and disjunctive desires allow desire theories of well-being to explain how a range of phenomena affect well-being. Moreover, these explanations enrich our understanding of the moral psychology of disappointment and pleasant surprises.

5.5 Chapter summary

I have argued that desire theories of well-being can explain why purported examples of Dead Sea apples do not improve well-being. They can do so by claiming that these examples do not describe genuine desire fulfilments. Rather, they describe cases where a frustrated conjunctive desire contains some true and some false propositions. This leaves us with the false impression that our desire has been fulfilled. Consequently, the problem of Dead Sea apples does not undermine the viability of desire theories of well-being. I then applied findings from this investigation to explain the effects of different types of pleasant surprises on well-being. I found that pleasant surprises can be explained through appeal to conjunctive and disjunctive desires. This allows desire theories of well-being to capture intuitions about the effects of pleasant surprises on well-being. Consequently, the existence of pleasant surprises poses no threat to the viability of these theories. It may be that these observations about the structure of desire can do additional work in advancing our understanding in both philosophy of well-being and moral psychology.

Conclusion

This thesis considered four main objections to desire theories of well-being. These are: the problem of self-sacrifice (Chapter Two), the problem of depression (Chapter Three), the problem of unstable desires (Chapter Four), and the problem of Dead Sea apples (Chapter Five). In each of these cases, I argued that desire theories of well-being have sufficient resources to solve these problems. In the course of doing so, I made a number of theoretical commitments. There are three that are central to the arguments made in this thesis. These are: that we have some basic standing desires and aversions for certain affective states (§1.3; §3.4; §4.3; §5.4), that desires do not always motivate proportionally to their strength (§2.4; §3.2; §3.4), and that some desires are conjunctive (§3.3; §5.3; §5.4). If any of these claims are incorrect, then several of my arguments will need to be reworked or rejected. Nevertheless, I am confident that these positions are independently plausible and defensible. I turn now to summarising each chapter of this thesis sequentially.

Chapter One opened with an overview of the literature on well-being. §1.1 outlined the concept of well-being and my approach to this investigation. §1.2 outlined a minimally plausible version of the desire theory of well-being and made the case for finding this view initially attractive. §1.3 examined mental state theories of well-being and gave reasons for rejecting these views in favour of desire theories of well-being. §1.4 repeated this approach with objective list theories. §1.5 concluded with a summary.

In §1.3 of this chapter, I committed to the view that human psychology is structured in such a way as to have some basic standing desires for certain positive affective states and some basic standing aversions to be free from certain negative affective states. This view does important work in several places in this thesis. In §3.4 I argued that to be sentient is to have at least some basic standing desires and aversions. This allowed me to claim that purported cases of complete conative collapse do not describe states of complete desirelessness. Consequently, they are not counterexamples to desire theories of well-being. In §4.3 I appealed to standing desires in order to explain why sometimes we are not benefitted by deliberately delaying the fulfilment of desires, and why fulfilling Ascending Desire is often more beneficial to us than fulfilling Descending Desire. And in §5.4 I appealed to standing desires as part of my explanation of how desire theories of well-being can capture the intuition that pleasant surprises improve well-being.

Chapter Two argued that desire theories of well-being are not undermined by the problem of self-sacrifice. This problem claims that these theories implausibly entail that self-sacrifice

does not exist. If this is right, then we ought to reject these theories. §2.1 outlined the problem of self-sacrifice. §2.2 surveyed existent proposed solutions to this problem and highlighted their shortcomings. §2.3 put forward a new solution to this problem based upon the rejection of proportionalism about desire and motivation. §2.4 highlighted independent reasons to reject proportionalism about desire and motivation. §2.5 concluded with a summary.

In §2.3 of this chapter, I committed to the view that desires do not always motivate proportionally to their strength. This claim is appealed to at several subsequent points in this thesis. In §3.2 I argued that severe depression suppresses the motivational force of some of our desires, while leaving the underlying desires intact. This allows desire theories of well-being to explain part of the harm of severe depression as emerging from the frustration of those desires that it leaves us unmotivated to fulfil. The claim that desires do not always motivate proportionally to their strength was also appealed to in §3.4 in order to explain how some desires to be badly off can motivate us with disproportional strength. Similar arguments were appealed to in §4.2 in order to claim that the strength of desires is not always proportional to the strength of their phenomenology.

Chapter Three argued that desire theories of well-being are not undermined by the problem of depression. This objection claims that these theories implausibly entail that severe depression does not lessen the well-being of those afflicted with this condition. If this is right, then we ought to reject these theories. §3.1 outlined the problem of depression. §3.2 argued that an effect of severe depression is that it decreases motivation while leaving some desires intact. §3.3 argued that another effect of severe depression is the inevitable frustration of desires that specify pleasure in their content. §3.4 examined residual cases of severe depression that are less well explained by these approaches. §3.5 considered the relationship between severe depression and prudential reasons. §3.6 concluded with a summary.

In §3.3 of this chapter, I appealed to the existence of conjunctive desires in order to explain one of the ways in which depression harms us. On this view, depression harms us by suppressing pleasure from emerging when pursuing our desires. I argue that many conjunctive desires specify both an object and that the subject with the desire takes pleasure in the existence of that object. Desires of this sort are inevitably frustrated when depression blocks pleasure from emerging. The existence of conjunctive desires is central to my proposed solution to the problem of Dead Sea apples in §5.3. Conjunctive desires also play a role in my explanation of the moral psychology of pleasant surprises in §5.4.

Chapter Four argued that desire theories of well-being are not undermined by the problem of unstable desires. This objection claims that these theories have implausible implications

about the relative importance to well-being of fleeting desires, long-standing desires, and fluctuations in desire strength. If this is right, then we ought to reject these theories. §4.1 outlined the problem of unstable desires. §4.2 argued that simple concurrentism cannot avoid this problem. §4.3 outlined stability-adjusted desire theories of well-being and defended these views from two objections. §4.4 considered value-fulfilment theories of well-being and argued that these views are less attractive solutions to the problem of unstable desires. §4.5 considered the implications of this chapter for our understanding of prudential reasons and certain mental disorders. §4.6 concluded with a summary.

Chapter Five argued that desire theories of well-being are not undermined by the problem of Dead Sea apples. This objection claims that these theories are implausible because they counterintuitively entail that desires which leave us disappointed and bereft of feelings of satisfaction nevertheless improve well-being. If this is right, then we ought to reject these theories. §5.1 outlined the problem of Dead Sea apples for desire theories of well-being. §5.2 surveyed existing approaches to this problem within the literature and argued that they are inadequate. §5.3 argued that reflections on the structure of desire allow us to explain Dead Sea apples as mere simulacra of desire fulfilments and consequently no threat to desire theories of well-being. §5.4 considered how similar observations can explain the effects of pleasant surprises on well-being. §5.5 concluded with a summary.

Throughout the course of this thesis, I have remained largely strategically ambiguous about which specific versions of the desire theory of well-being are most plausible. Consequently, readers with different opinions about how these theories ought to be constructed should be able to endorse large parts of this thesis. The exception to this strategic ambiguity is found in Chapter Four. There I argued that any plausible version of this view will need to be stability-adjusted. Aside from this, in §5.2.1 I expressed sympathy with versions of the desire theory of well-being that reject concurrentism. And in §5.2.3 I expressed sympathy with the view that only the fulfilment and frustration of intrinsic desires count towards well-being. My discussion of the Humean Theory of Motivation in Chapter Two may also betray some sympathy towards that position. Nevertheless, none of the central arguments of this thesis require specific positions on these issues. Consequently, I remain tentative in my support for these views.

Desire theories of well-being are relatively popular within analytic philosophy. A 2020 survey of philosophers' attitudes towards a range of different philosophical positions found that 18.6% of us accepted or leant towards accepting views of this sort (Bourget & Chalmers 2023, 14). If this thesis is successful, then it will have deprived opponents of these views of

several influential arguments against them. It will also have expanded the range of conceptual resources available to those who wish to defend these views. A secondary set of findings relate to advancing our understanding of moral psychology more broadly. It is possible that these findings may have implications in fields beyond philosophy of well-being.

List of References

Adams, R. M. (1999). *Finite and infinite goods: A framework for ethics*. New York: Oxford University Press.

Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford: Oxford University Press.

Almeder, R. F. (2000). Human happiness and morality: A brief introduction to ethics. Amherst, NY: Prometheus Books.

Archer, A. (2016). Supererogation, sacrifice, and the limits of duty. *The Southern Journal of Philosophy*, 54(3), 333–354.

Aristotle (2014). *Aristotle's Ethics: Writings from the complete works*. J. Barnes & A. Kenny (eds). Oxford: Princeton University Press.

Arpaly, N. & Schroeder, T. (2014). *In praise of desire*. New York: Oxford University Press.

Baber, H. E. (2010). Ex ante desire and post hoc satisfaction. In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (eds.), *Time and identity* (pp. 249–267). Cambridge, MA: MIT Press.

Barry, B. (1989). Utilitarianism and preference change. *Utilitas*, 1(2), 278–282.

Bentham, J. (1964) [1789]. An introduction to the principles of morals and legislation. In L. A. Selby-Bigge (ed.), *British moralists*. Indianapolis, IN & New York: Bobbs-Merrill.

Bourget, D. & D. Chalmers (2023). Philosophers on philosophy: The 2020 PhilPapers survey. *Philosophers' Imprint*, 23(11), 1–53.

Bradley, B. (2007). A paradox for some theories of welfare. *Philosophical Studies*, 133(1), 45–53.

Bradley, B. (2008). The worst time to die. *Ethics*, 118(2), 291–314.

Bramble, B. (2013). The distinctive feeling theory of pleasure. *Philosophical Studies*, 162(2), 201–217.

Brandt, R. (1972). Rationality, egoism, and morality. *Journal of Philosophy*, 69(20), 681–697.

Brandt, R. (1979). *A theory of the good and the right*. Oxford: Clarendon Press.

Brewer, T. (2006). Three dogmas of desire. In T. Chappell (ed.), *Values and virtues: Aristotelianism in contemporary ethics* (pp. 253–285). Oxford: Clarendon Press.

Bruckner, D. W. (2013). Present desire satisfaction and past well-being. *Australasian Journal of Philosophy*, 91(1), 15–29.

Bruckner, D. W. (2016). Quirky desires and well-being. *Journal of Ethics and Social Philosophy*, 10(2), 1–34.

Bykvist, K. (2003). The moral relevance of past preferences. In H. Dyke (ed.), *Time and ethics: Essays at the intersection* (pp. 115–136). Dordrecht: Kluwer Academic Publishers.

Carson, T. (2000). *Value and the good life*. Notre Dame: University of Notre Dame Press.

Chang, R. (1997). Introduction. In R. Chang (ed.), *Incommensurability, incomparability, and practical reason* (pp. 1–34). Cambridge, MA & London: Harvard University Press.

Crisp, R. (2006a). Hedonism reconsidered. *Philosophy and Phenomenological Research*, 73(3), 619–645.

Crisp, R. (2006b). *Reasons and the good*. New York: Oxford University Press.

Dancy, J. (1993). *Moral reasons*. Oxford: Basil Blackwell.

Dancy, J. (2000). *Practical reality*. Oxford: Oxford University Press.

Darwall, S. (2002). *Welfare and rational care*. Princeton, NJ: Princeton University Press.

Dietz, A. (2023). Making desires satisfied, making satisfied desires. *Philosophical Studies*, 180(3), 979–999.

Donner, W. (1991). *The liberal self: John Stuart Mill's moral and political theory*. New York: Cornell University Press.

Dorsey, D. (2012a). Intrinsic value and the supervenience principle. *Philosophical Studies*, 157(2), 267–285.

Dorsey, D. (2012b). Subjectivism without desire. *The Philosophical Review*, 121(3), 407–442.

Dorsey, D. (2013). Desire-satisfaction and welfare as temporal. *Ethical Theory and Moral Practice*, 16(1), 151–171.

Dorsey, D. (2017). Idealization and the heart of subjectivism. *Noûs*, 51(1), 196–217.

Dorsey, D. (2019). Preferences and prudential reasons. *Utilitas*, 31(2), 157–178.

Dorsey, D. (2021). *A theory of prudence*. Oxford: Oxford University Press.

Elster, J. (1983). *Sour grapes: Studies in the subversion of rationality*. New York: Cambridge University Press.

Enoch, D. (2005). Why idealize? *Ethics*, 115(4), 759–787.

Fara, D. G. (2013). Specifying desires. *Noûs*, 47(2), 250–272.

Fehige, C. (1998). A Pareto Principle for possible people. In C. Fehige & U. Wessels (eds.), *Preferences* (pp. 508–540). Berlin & New York: Walter de Gruyter.

Feldman, F. (1991). Some puzzles about the evil of death. *The Philosophical Review*, 100(2), 205–227.

Feldman, F. (2002). The good life: A defense of attitudinal hedonism. *Philosophy and Phenomenological Research*, 65(3), 604–628.

Feldman, F. (2004). *Pleasure and the good life: Concerning the nature, varieties, and plausibility of hedonism*. Oxford: Oxford University Press.

Firth, R. (1952). Ethical absolutism and the ideal observer. *Philosophy and Phenomenological Research*, 12(3), 317–345.

Fletcher, G. (2013). A fresh start for the objective-list theory of well-being. *Utilitas*, 25(2), 206–220.

Fletcher, G. (2016). *The philosophy of well-being: An introduction*. Oxon: Routledge.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5–20.

Frankfurt, H. (2004). *The reasons of love*. Princeton, NJ: Princeton University Press.

Frankl, V. (1986). *The doctor and the soul*. New York: Random House.

Goldman, A. H. (2009). *Reasons from within: Desires and values*. New York: Oxford University Press.

Gregory, A. (2012). Changing direction on direction of fit. *Ethical Theory and Moral Practice*, 15(5), 603–614.

Gregory, A. (2021). *Desire as belief: A study of desire, motivation, and rationality*. Oxford: Oxford University Press.

Griffin, J. (1986). *Well-being: Its meaning, measurement, and moral importance*. Oxford: Clarendon Press.

Hardie, W. F. R. (1965). The final good in Aristotle's ethics. *Philosophy*, 40(154), 277–295.

Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. Oxford: Clarendon Press.

Hari, J. (2018). *Lost connections: Uncovering the real causes of depression – And the unexpected solutions*. London: Bloomsbury.

Heathwood, C. (2005). The problem of defective desires. *Australasian Journal of Philosophy*, 83(4), 487–504.

Heathwood, C. (2006). Desire satisfactionism and hedonism. *Philosophical Studies*, 128(3), 539–563.

Heathwood, C. (2011). Preferentism and self-sacrifice. *Pacific Philosophical Quarterly*, 92(1), 18–38.

Heathwood, C. (2015). Desire-fulfillment theory. In G. Fletcher (ed.), *The Routledge handbook of philosophy of well-being* (pp. 135–147). London & New York: Routledge.

Heathwood, C. (2019). Which desires are relevant to well-being? *Noûs*, 53(3), 664–688.

Heinaman, R. (1993). Eudaimonia and Kakodaimonia in Aristotle. *Phronesis*, 38(1), 31–56.

Hubin, D. C. (1996). Hypothetical motivation. *Noûs*, 30(1), 31–54.

Hubin, D. C. (2003). Desires, whims and values. *The Journal of Ethics*, 7(3), 315–335.

Kagan, S. (1994). Me and my life. *Proceedings of the Aristotelian Society, New Series*, 94(1), 309–324.

Kagan, S. (2014). An Introduction to ill-being. In M. Timmons (ed.), *Oxford studies in normative ethics, volume 4* (pp. 261–288). Oxford: Oxford University Press.

Keller, S. (2009). Welfare as success. *Noûs*, 43(4), 656–683.

Kelly, C. (2008). The impossibility of incommensurable values. *Philosophical Studies*, 137(3), 369–382.

Kendler, K. S., Myers, J. & Halberstadt, L. J. (2010). Should the diagnosis of major depression be made independent of or dependent upon the psychosocial context? *Psychological Medicine*, 40, 771–780.

Khader, S. J. (2011). *Adaptive preferences and women's empowerment*. New York: Oxford University Press.

Lauinger, W. (2011). Dead sea apples and desire-fulfillment welfare theories. *Utilitas*, 23(3), 324–343.

Leichsenring, F., Heim, N., Leweke, F., Spitzer, C., Steinert C. & Kernberg, O. F. (2023). Borderline personality disorder: A review. *JAMA*, 329(8), 670–679.

Lewis, D. (2000). Dispositional theories of value. In D. Lewis (ed.), *Papers in ethics and social philosophy* (pp. 68–94). New York: Cambridge University Press.

Lin, E. (2017a). Against welfare subjectivism. *Noûs*, 51(2), 354–377.

Lin, E. (2017b). Asymmetrism about desire satisfactionism and time. In M. Timmons (ed.), *Oxford studies in normative ethics, volume 7* (pp. 161–183). Oxford: Oxford University Press.

Lin, E. (2018). Welfare invariabilism. *Ethics*, 128(2), 320–345.

Loeb, D. (1995). Full-information theories of individual good. *Social Theory and Practice*, 21(1), 1–30.

Lowe, D. & Stenberg, J. (2017). The experience machine objection to desire satisfactionism. *Journal of the American Philosophical Association*, 3(2), 247–263.

Lukas, M. (2010). Desire satisfactionism and the problem of irrelevant desires. *Journal of Ethics and Social Philosophy*, 4(2), 1–25.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. London: Penguin Books.

Mariqueo-Russell, A. (2023a). Desire and motivation in desire theories of well-being. *Philosophical Studies*, 180(7), 1975–1994.

Mariqueo-Russell, A. (2023b). Well-being and the problem of unstable desires. *Utilitas*, 35(4), 260–276.

McDaniel, K. & Bradley, B. (2008). Desires. *Mind, New Series*, 117(466), 267–302.

Mendola, J. (2009). Real desires and well-being. *Philosophical Issues*, 19(1), 148–165.

Mill, J. S. (1998) [1861]. *Utilitarianism*. Oxford: Oxford University Press.

Milton, J. (1667). *Paradise Lost*. London: Samuel Simmons.

Murphy, M. (1999). The simple desire-fulfillment theory. *Noûs*, 33(2), 247–272.

Nagel, T. (1970a). *The possibility of altruism*. Princeton, NJ: Princeton University Press.

Nagel, T. (1970b). Death. *Noûs*, 4(1), 73–80.

Noggle, R. (1999). Integrity, the self, and desire-based accounts of the good. *Philosophical Studies*, 96(3), 301–328.

Nozick, R. (1974). *Anarchy, state, and utopia*. Malden, MA, Oxford & Carlton, OR: Blackwell Publishing.

Nussbaum, M. (1986). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge: Cambridge University Press.

Nussbaum, M. (2016). *Anger and forgiveness: Resentment, generosity, justice*. New York: Oxford University Press.

Overvold, M. C. (1980). Self-interest and the concept of self-sacrifice. *Canadian Journal of Philosophy*, 10(1), 105–118.

Pallies, D. (2022). Attraction, aversion, and asymmetrical desires. *Ethics*, 132(3), 598–620.

Parfit, D. (1984). *Reasons and persons*. New York: Oxford University Press.

Penelhum, T. (1979). Human nature and external desires. *The Monist*, 62(3), 304–319.

Pineda-Oliva, D. (2021). Defending the motivational theory of desire. *Theoria*, 36(2), 243–260.

Pitcher, G. (1984). The misfortunes of the dead. *American Philosophy Quarterly*, 21(2), 183–188.

Porter, E. (2023). *Autonomy as an ideal for neuro-atypical agency: Lessons from bipolar disorder*. PhD thesis. University of Kent.

Portmore, D. W. (2007). Desire fulfillment and posthumous harm. *American Philosophical Quarterly*, 44(1), 27–38.

Raibley, J. (2010). Well-being and the priority of values. *Social Theory and Practice*, 36(4), 593–620.

Railton, P. (1986). Facts and values. *Philosophical Topics*, 14(2), 5–31.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Belknap Press.

Richardson, H. (1994). *Practical reasoning about final ends*. Cambridge: Cambridge University Press.

Rosati, C. S. (1995). Persons, perspectives, and full information accounts of the good. *Ethics*, 105(2), 296–325.

Rosati, C. S. (1996). Internalism and the good for a person. *Ethics*, 106(2), 297–326.

Rosati, C. S. (2009) Self-interest and self-sacrifice. *Proceedings of the Aristotelian Society, New Series*, 109(1pt3), 311–325.

Sarch, A. (2013). Desire satisfaction and time. *Utilitas*, 25(2), 221–245.

Schoeman, F. D. (1978). Responsibility and the problem of induced desires. *Philosophical Studies*, 34(3), 293–301.

Schroeder, M. (2007). *Slaves of the passions*. Oxford: Oxford University Press.

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

Schueler, G. F. (1995). *Desire: Its role in practical reason and the explanation of action*. Cambridge, MA: MIT Press.

Sen, A. (1984). *Resources, values and development*. Oxford: Basil Blackwell.

Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford: Clarendon Press.

Sidgwick, H. (1981) [1907]. *The methods of ethics*. 7th ed. Indianapolis, IN: Hackett Publishing Company.

Silverstein, M. (2000). In defense of happiness: A response to the experience machine. *Social Theory and Practice*, 26(2), 279–300.

Sinhababu, N. (2015). Advantages of propositionalism. *Pacific Philosophical Quarterly*, 96(2), 165–180.

Sinhababu, N. (2017). *Humean Nature: How desire explains action, thought, and feeling*. Oxford: Oxford University Press.

Skow, B. (2009). Preferentism and the paradox of desire. *Journal of Ethics and Social Philosophy*, 3(3), 1–16.

Smilansky, S. (2021). A puzzle about self-sacrificing altruism. *Journal of Controversial Ideas*, 1(1), 1–14.

Smith, M. (1994). *The moral problem*. Oxford: Blackwell Publishing.

Smuts, A. (2011). The feels good theory of pleasure. *Philosophical Studies*, 155(2), 241–265.

Sobel, D. (1997). On the subjectivity of welfare. *Ethics*, 107(3), 501–508.

Sobel, D. (1998). Well-being as the object of moral consideration. *Economics and Philosophy*, 14(2), 249–281.

Sobel, D. (2009). Subjectivism and idealization. *Ethics*, 119(2), 336–352.

Spaid, A. (2020). *Desire satisfaction theories and the problem of depression*. PhD thesis. University of Nebraska.

Strawson, G. (2010). *Mental reality*. 2nd ed. Cambridge, MA & London: MIT Press.

Stocker, M. (1979). Desiring the bad: An essay in moral psychology. *Journal of Philosophy*, 76(12), 738–753.

Sumner, L. W. (1996). *Welfare, happiness, and ethics*. Oxford: Clarendon Press.

Sunstein, C. R. (2018). *The cost-benefit revolution*. Cambridge, MA & London: The MIT Press.

Swartzer, S. (2015). Humean externalism and the argument from depression. *Journal of Ethics and Social Philosophy*, 9(2), 1–16.

Thagard, P. (2006). Desires are not propositional attitudes. *Dialogue*, 45(1), 151–156.

Tolstoy, L. (1954). *A confession, the gospel in brief, and what I believe*. London: Oxford University Press.

Tully, I. (2017). Depression and the problem of absent desires. *Journal of Ethics and Social Philosophy*, 11(2), 1–15.

Tully, I. (2019). Demarcating depression. *Ratio*, 32(2), 114–121.

Vadas, M. (1984). Affective and non-affective desire. *Philosophy and Phenomenological Research*, 45(2), 273–279.

van der Deijl, W. (2019). Is pleasure all that is good about experience? *Philosophical Studies*, 176(7), 1769–1787.

van Weelden, J. (2019). On two interpretations of the desire-satisfaction theory of prudential value. *Utilitas*, 31(2), 137–156.

Velleman, D. (1991). Well-being and time. *Pacific Philosophical Quarterly*, 72(1), 48–77.

Vorobej, M. (1998). Past desires. *Philosophical Studies*, 90(3), 305–318.

Wilkes, K. V. (1978). The good man and the good for man in Aristotle's ethics. *Mind, New Series*, 87(348), 553–571.

Williams, B. (1973). Egoism and altruism. In B. Williams (ed.), *Problems of the self: Philosophical papers 1956–1972*, (pp. 250–265). Cambridge: Cambridge University Press.

Williams, E. G. (2016). Preferences' significance does not depend on their content. *Journal of Moral Philosophy*, 13(2), 211–234.

Wolf, S. (1982). Moral saints. *Journal of Philosophy*, 79(8), 419–439.

Wolf, S. (1997). Happiness and meaning: Two aspects of the good life. *Social Philosophy and Policy*, 14(1), 207–225.

Woodard, C. (2013). Classifying theories of welfare. *Philosophical Studies*, 165(3), 787–803.

Yelle, B. (2014). Alienation, deprivation, and the well-being of persons. *Utilitas*, 26(4), 367–384.

Yelle, B. (2016). In defense of sophisticated theories of welfare. *Philosophia*, 44(4), 1409–1418.

Yu, X. (2022). Hidden desires: A unified strategy for defending the desire-satisfaction theory. *Utilitas*, 34(4), 445–460.