

HUMAN WELL-BEING:  
THE NO PRIORITY THEORY

A Dissertation  
submitted to the Faculty of the  
Graduate School of Arts and Sciences  
of Georgetown University  
in partial fulfillment of the requirements for the  
degree of  
Doctor of Philosophy  
in Philosophy

By

William A. Lauinger

Washington, D.C.  
June 22, 2009

UMI Number: 3371617

### INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.



UMI Microform 3371617  
Copyright 2009 by ProQuest LLC  
All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

Copyright 2009 by William A. Lauinger  
All Rights Reserved

HUMAN WELL-BEING:  
THE NO PRIORITY THEORY

William A. Lauinger

Thesis Advisor: Mark Murphy, Ph.D.

ABSTRACT

Desire-fulfillment (DF) theories and objective list (OL) theories are the two dominant types of theories of human well-being. I argue that DF theories fail to capture the *good* part of ‘good for’, and that OL theories fail to capture the *for* part of ‘good for’. Then, with the aim of capturing both of these parts of ‘good for’, I advance a theory that places both a value constraint and a desire constraint on human well-being. The most important objection to my theory comes from OL theorists. My theory implies that, if one lacks the desire for accomplishment, or friendship, or health, etc., then this non-desired good isn’t part of one’s welfare. OL theorists find this implication of my theory unacceptable. However, I argue that there is no genuine worry for my theory here, because (1) the vast majority of humans just do desire accomplishment, friendship, health, etc., and stably so, and because (2) in those very rare cases where the desires for these goods are missing, we don’t have sufficient reasons for thinking that the relevant non-desired goods really are part of the welfare of the agents in question.

## ACKNOWLEDGEMENTS

I should thank just about all of the professors in the Georgetown philosophy department, since just about all of them have helped me along the way. Special thanks to Wilfried Ver Eecke, John Langan, Alex Pruss, and Mark Murphy, since they have helped me the most. I should single out Mark Murphy here, because he has been an excellent chair for me: always helpful (and promptly so), always clear and thorough in his guidance, always demanding (as a chair should be), and always good-spirited (even when he disagreed with me!). I count myself very fortunate to have had him as my chair.

Thanks also go to the other graduate students, both the ones who are now gone and the ones who are still around. David deserves special thanks. Our talks – usually about either basketball or philosophy – have always made me very happy.

As for family members, I should thank my parents, as well as my brothers and sisters and their families. Special thanks to Katie here, since she is the one who really got me into philosophy in the first place. Most of all, I should thank my wife, Kym, and my daughter, Maddy. They are the best, and this dissertation is dedicated to them. Lastly, I should make sure to thank our three-legged chihuahua, Baby, who weighs a grand total of 2.5 pounds. She sat next to me in her little box as I typed this dissertation. Amazingly, she did not object to anything that I wrote.

## TABLE OF CONTENTS

<b>Introduction</b> .....	1
1. The problem.....	1
2. The proposed solution: no priority.....	3
3. Hybrid theories in general.....	8
4. The near-possible versus far-possible distinction.....	9
5. Desires.....	12
<b>Chapter One: Desire-fulfillment Theories</b> .....	23
6. What might be attractive about DF theories?.....	23
7. DF theories and the stability problem.....	27
8. DF theories and sufficient broadness.....	29
9. DF theories and sufficient narrowness: four objections.....	34
10. DF theories and sufficient narrowness: a fifth objection.....	44
11. Second-order DF theories.....	58
12. How do hypothetical DF theories work?.....	64
13. Hypothetical DF theories and the stability problem.....	68
14. Hypothetical DF theories and the <i>good</i> part of ‘good for’.....	74
15. Hypothetical DF theories and the <i>for</i> part of ‘good for’.....	85
<b>Chapter Two: Objective List Theories</b> .....	92
16. Two common objectivist moves.....	92
17. The main strengths of OL theories.....	96
18. Non-standard OL theories and prudential internalism.....	100

19. Prudential internalism: what it says and doesn't say, and why it seems true....	105
20. Prudential internalism: interpersonal justification, intrapersonal deliberation...	119
21. OL theories and prudential internalism: why the hypothetical route fails.....	127
<b>Chapter Three: The No Priority Theory.....</b>	<b>137</b>
22. The no priority theory: comments on its value and desire conditions.....	137
23. The main objection to the no priority theory: the missing desires objection.....	141
24. Standard cases.....	144
25. Non-standard cases not involving psychological disorders.....	153
26. Non-standard cases: depression, eating disorders, and masochism.....	158
27. Non-standard cases: autism.....	169
28. Non-standard cases: psychopathy.....	175
29. The no priority theory and babies' desires.....	183
30. The relationship between prudential value and perfectionist value.....	185
<b>Chapter Four: Hybrid Theories in General.....</b>	<b>195</b>
31. Brief remarks on the main hybrid theory options.....	195
32. Should we rely on enjoyment?.....	196
33. Should we rely on cognitive endorsement?.....	203
34. How should a hybrid theory that relies on desire be formulated?.....	208
35. Closing comments.....	211
<b>Bibliography.....</b>	<b>213</b>
<b>Notes.....</b>	<b>217</b>

## Introduction

### 1. The problem

Desire-fulfillment theories (DF theories) and objective list theories (OL theories) are currently the two most commonly defended types of theories of human welfare. DF theories state that something is an aspect (or component) of someone's welfare if and only if, and directly because, she desires it.<sup>1</sup> OL theories state that something is an aspect of someone's welfare if and only if, and directly because, it instantiates one of the basic goods for her, where the basic goods are things such as friendship, accomplishment, knowledge, and aesthetic experience.

OL theorists often contend that DF theories cannot capture the *good* part of 'good for', and DF theorists often contend that OL theories cannot capture the *for* part of 'good for'. Without yet entering into any details, it will help if I give us a brief sense of how this debate runs, for doing so will help us in setting up the problem that my welfare theory – call it *the no priority theory* – is meant to solve.

OL theorists commonly criticize DF theories by saying the following. 'DF theories are inadequate in at least two respects. (1) They can't account for the fact that we can desire things that aren't aspects of our welfare. If someone has a crazy desire – say, the desire to eat glass – then it follows on DF theories that his eating glass is an aspect of his welfare. But that seems wrong, since, barring strange circumstances, it seems that eating glass can't be intrinsically *good* for someone. (2) DF theories can't handle the fact that we can fail to desire things that are aspects of our welfare. If someone lacks the desire for friends, then it follows on DF theories that having friends isn't part of his welfare. And that seems wrong – or, at any rate, not quite right.'



DF theorists often criticize OL theories by saying the following. ‘In speaking of someone’s well-being, we aren’t (or aren’t just) speaking of what is intrinsically good – we are speaking of what is intrinsically good *for her*. Thus welfare (or prudential value) is a *relational* sort of value; and, more than this, it seems to be an especially *personal* sort of value (since, after all, someone’s well-being is particularly suited for her).<sup>2</sup> In order to capture the distinctively relational and personal nature of welfare, it seems that a welfare theory must embrace a relatively strong form of *prudential internalism*. Taking X to be a state of affairs and A to be a human, prudential internalism says that X can’t be an aspect of A’s welfare unless there is some sort of a positive hookup between X and A.<sup>3</sup> The basic intuition here is this: It can’t be true that X is *both* intrinsically good for A (or an aspect of A’s welfare) *and* thoroughly alienating to A. In its most straightforward form prudential internalism says that, if X is an aspect of A’s welfare, then A must have some sort of pro-attitude toward X. It seems that OL theories cannot accept any relatively strong form of prudential internalism. Suppose that X instantiates, say, Doug’s having friendship. Here it follows on OL theories that X is an aspect of Doug’s welfare; and this follows on OL theories *even if Doug has no pro-attitude of any sort toward X*. Cases such as this one show that OL theories can’t provide us with a sufficiently non-alienating account of human well-being.’

To sum up, then, the important charges that are being leveled here are these: DF theories can’t guarantee the *good* part of ‘good for’, and OL theories can’t guarantee the *for* part of ‘good for’. This sets up a problem for us, namely, the problem of capturing both the *good* and the *for* parts of ‘good for’. We will be grappling with this problem – call it *the prudential problem* – throughout all of what follows.

In dealing with the prudential problem, one approach is to stick with the DF theory structure or the OL theory structure and then to fill out and/or amend and/or qualify until we have a theory that seems to capture both the *good* and the *for* parts of ‘good for’ in a sufficient way. I will argue that this approach fails. In Chapter 1 we will examine DF theories in detail, and in Chapter 2 we will examine OL theories in detail; and, with respect to both DF theories and OL theories, my argument will be that they cannot guarantee both the *good* and the *for* parts of ‘good for’.

## **2. The proposed solution: no priority**

An obvious solution to the prudential problem, and my own favored solution, is to claim that human well-being is a function *both* of goodness that is independent of desire *and* of desire. The no priority theory is built upon this idea: The no priority theory claims that X is an aspect of A’s well-being if and only if, and directly because, (1) X is intrinsically valuable in the appropriate desire-independent way and (2) X’s obtaining satisfies some desire that A has (where this should be understood in a certain way). I will elaborate on each of these two conditions – the value condition and the desire condition – in Chapter 3. It will help, though, if I provide a brief, introductory discussion of each of these two conditions here.

*The value condition:* The no priority theory states that, in order for X to be an aspect of A’s well-being, X must be intrinsically valuable in the appropriate desire-independent way. But what sort of intrinsic value do I have in mind here? I don’t have *prudential* value in mind. After all, it would be problematically circular if I were to try to explain prudential value (or welfare) partly in terms of prudential value (or welfare).

Also, I don't have any sort of *non-relational* intrinsic value in mind. Suppose that a natural disaster (say, a tornado) strikes a town and victimizes many people there; and suppose that you form the desire that these people recover from their misfortune (perhaps you even give a small donation). And now suppose that they do end up recovering. Their recovering seems not only to be good for them, but also to be non-relationally and intrinsically good. However, their recovering seems to be *too independent of you* to touch upon (or to enter) your welfare; and this seems especially to be true if you aren't even aware of their recovering.<sup>4</sup> In order to avoid having to worry about cases such as this one, it is best to formulate the no priority theory not in terms of some sort of non-relational intrinsic value, but rather in terms of some sort of *relational* intrinsic value. This is a tricky matter, though, for I need this sort of value not only to be relational and intrinsic, but also to be *desire-independent*. Here I will invoke the same goods that OL theorists invoke: friendship, accomplishment, aesthetic experience, health, etc. However, I won't conceive of these goods – following OL theorists, let's call them the basic goods – in the same way that OL theorists conceive of them. OL theorists conceive of these goods as being the fundamental forms of *prudential* value. But I will conceive of these goods as being the fundamental forms of some *non-prudential* sort of value – that is, some non-prudential sort of value that is relational, intrinsic, and desire-independent.

If I must specify the sort of value in question here (that is, beyond saying that it is non-prudential, relational, intrinsic, and desire-independent), then I will claim that it is *perfectionist* value. Something (anything) is a perfectionist good for someone if and only if (and because) it directly contributes to the proper/excellent functioning of her *qua* her being the sort of thing she is, namely, a human being. My idea here, then, is this. We

can take the basic goods (accomplishment, knowledge, friendship, etc.) and then claim that we should conceive of them as being perfectionist goods. Moreover, we can claim that, if X is, for A, an instance of friendship (or health, or knowledge, or whatever), then X is intrinsically good for A in the perfectionist sense (and X is so independently of A's desiring/not desiring X).

Perfectionist value does seem to be a candidate for the appropriate sort of value here, for it does seem to be a non-prudential sort of value that is relational, intrinsic, and desire-independent. Think of the matter this way. Perfectionist value is like prudential value in that it is *relational*: Perfectionist value is always *for* someone or something in that it is always at least pro tanto perfective of someone or something. But, in spite of its being relational, perfectionist value doesn't seem to be *personal* in the same way that prudential value is. This point is hard to get right. But, very roughly, we can say that, whereas prudential value is intrinsically good for someone *qua* her being the *individual* human that she is, perfectionist value seems instead to be intrinsically good for someone *qua* her being a human *as such*. Given, then, that perfectionist value isn't *personal* in the same way that prudential value is, it is sensible to think that perfectionist value isn't desire-dependent in the same way that prudential value is; that is, it is sensible to think that, although what is intrinsically good for one in the prudential sense is at least partly dependent on one's own desires, what is intrinsically good for one in the perfectionist sense isn't at all dependent on one's own desires.

*The desire condition*: I have in mind *actual* desires, not hypothetical desires (or the desires that one would have, say, if one were better informed than one actually is). Also, I have in mind *intrinsic* desires, not merely instrumental desires. (Unless I note

otherwise, I always use the term ‘desire’ to refer to intrinsic desires.) Further, my claim isn’t that A must desire X in order for X to count as an aspect of A’s welfare. Rather, my claim is that, in order for X to be an aspect of A’s welfare, X’s obtaining must satisfy some desire that A has. Suppose that I desire aesthetic experience, and also suppose that hearing a certain song instantiates aesthetic experience for me. Even if I have no antecedent desire to hear this song, my hearing this song nonetheless satisfies a desire that I antecedently have, namely, the desire for aesthetic experience. (Here, if we want, we can use ‘pro tanto satisfies’ instead of ‘satisfies’. Using ‘pro tanto satisfies’ rather than ‘satisfies’ would suggest that my desire for aesthetic experience isn’t *completely* satisfied by my hearing this song. In other words, using ‘pro tanto satisfies’ rather than ‘satisfies’ would suggest that my desire for aesthetic experience is only *to an extent* satisfied by my hearing this song. For the sake of brevity/simplicity, I will always use ‘satisfies’ – that is, I will never use ‘pro tanto satisfies’ – when referring to the desire satisfactions that matter on the no priority theory.)

In light of the above remarks, it should be clear what the no priority theory says. It says that X is an aspect of A’s welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, of health, or friendship, or knowledge, etc.) and (2) X’s obtaining satisfies some desire that A has (with the idea being that, if A does not desire X itself, then A must at least desire the basic good that X here instantiates for A).

We might now consider whether the no priority theory solves the prudential problem: the problem of guaranteeing both the *good* and the *for* parts of ‘good for’. At least at first glance, the answer appears to be yes. After all, the value condition seems to

guarantee the *good* part of ‘good for’, while the desire condition seems to guarantee the *for* part of ‘good for’. There is, however, an important objection that we should consider. OL theorists commonly claim that DF theories cannot handle the fact that someone can fail to desire various things that look for all the world to be prudentially good for her (say, things such as accomplishment, friendship, aesthetic experience, and health). And this same objection, suitably altered, can be – and no doubt will be – leveled against the no priority theory.<sup>5</sup> As applied to my theory (the no priority theory), the objection in question – call it *the missing desires objection* – runs at least roughly as follows. My theory entails that goods such as accomplishment, knowledge, health, and aesthetic experience are part of one’s welfare only if, and only when, one happens to desire them. And that seems wrong, for in truth these goods seem to be stable parts of one’s welfare. Indeed, each of these goods seems to be included within one’s welfare even if the desire for it happens to be missing. Or, put differently, the objection here is this: My theory entails that A’s welfare is partly a function of A’s desires, but that is implausible, for A’s desires are contingent occurrences that may or may not happen to track various goods (friendship, health, knowledge, etc.) that seem rather stably to be part of A’s welfare. It is best, then, to reject my theory and instead to claim that human welfare is a desire-independent sort of value.

In Chapter 3 we will discuss the no priority theory in detail. There I will argue that the no priority theory can handle the missing desires objection because (1) the vast majority of humans just do desire accomplishment, friendship, health, etc., and stably so, and because (2) in those very rare cases where the desires for these goods are missing, we don’t have sufficient reasons for thinking that the relevant non-desired goods really are

part of the welfare of the agents in question. (As a shorthand way of saying that the vast majority of humans just do desire friendship, health, aesthetic experience, knowledge, etc., and stably so, I will often say that human desire with respect to the basic goods is remarkably widespread and stable.)

### **3. Hybrid theories in general**

The no priority theory is one type of hybrid theory. Many philosophers have recently advanced theories that are at least in the ballpark of being hybrid theories. Here, for instance, I have in mind Robert Adams (who speaks of one's welfare as primarily consisting in the enjoyment of the excellent), Stephen Darwall (who speaks of one's welfare as primarily consisting in the appreciation of agent-neutral goods), and Connie Rosati (who speaks of one's welfare as consisting in a "satisfying engagement with objective values").<sup>6</sup> All hybrid theories claim that human welfare is a function *both* of some intrinsic, non-prudential, and pro-attitude independent sort of value *and* of some sort of pro-attitude. Different hybrid theorists rely on different sorts of non-prudential value. I rely on a *relational* sort of non-prudential value, but one might rely on a *non-relational* sort of non-prudential value. Also, different hybrid theorists rely on different pro-attitudes. I rely on *desire*, but one might rely on, say, *enjoyment*. There is, however, a further matter to consider here. How widespread, and how stable, are the pro-attitudes that partly constitute well-being?

The no priority theory as such says nothing about this matter. But, in defending the no priority theory against the missing desires objection, it is important to argue that human desire with respect to the basic goods is remarkably widespread and stable. To

my knowledge, there are no contemporary hybrid theorists (except for me) that place any explicit emphasis on the remarkably widespread presence of, and the remarkable stability of, the pro-attitudes that partly constitute well-being. And that, I think, is a major problem, for it seems that no hybrid theory can be adequately defended against missing pro-attitude sorts of objections unless an explicit emphasis is placed on the remarkably widespread presence of, and the remarkable stability of, the pro-attitudes that partly constitute well-being.

I will discuss the details of hybrid theories other than the no priority theory in Chapter 4. Until then, though, I will leave these other hybrid theories to the side.

#### **4. The near-possible versus far-possible distinction**

We should distinguish between (what I will refer to as) near-possible and far-possible aspects of welfare. In understanding this distinction it helps to think about the question of *when* states of affairs become prudentially valuable. Here we might consider DF theories and OL theories. On DF theories states of affairs become prudentially good for oneself the moment that one desires them. It doesn't matter whether the desired states of affairs obtain or not. If one now desires them, then they are now prudentially good for oneself. Thus, if Nikhil now desires the state of affairs *Nikhil's drinking/metabolizing milk*, then this state of affairs is now prudentially good for Nikhil (and it is so whether it obtains or not). The same sort of thing holds on OL theories, though for a different reason. On OL theories states of affairs become prudentially good for oneself the moment that (vis-à-vis oneself) they instantiate (or fall under) the basic goods. Indeed, if (vis-à-vis oneself) given states of affairs now fall under the basic goods, then such states



of affairs are now prudentially good for oneself (and they are so whether they obtain or not). Thus, if the state of affairs *Nikhil's drinking/metabolizing milk* now falls under the basic good of health for Nikhil, then this state of affairs is now prudentially good for Nikhil (and it is so whether it obtains or not).

In this way, then, we are led to the distinction between near-possible and far-possible aspects of welfare. 'X is a near-possible aspect of A's welfare' should be taken to mean 'X is now prudentially good for A'. The near-possible aspects of one's welfare fall into two types: the merely near-possible and the actual types. 'X is a merely near-possible aspect of A's welfare' should be taken to mean 'X is now prudentially good for A, but X doesn't now obtain'. 'X is an actual aspect of A's welfare' should be taken to mean 'X is now prudentially good for A, and X now obtains'. As for far-possible aspects of welfare, 'X is a far-possible aspect of A's welfare' should be taken to mean 'X isn't now prudentially good for A, but X might later come to be prudentially good for A'.

When I speak of aspects of welfare, I typically just use 'X is an aspect of A's welfare'. But by 'X is an aspect of A's welfare' (or by 'X is prudentially good for A') I always mean 'X is a *near-possible* aspect of A's welfare'. Whenever I am speaking solely of *actual* aspects of welfare, I will be explicit about my doing so. And the same goes for *merely near-possible* and *far-possible* aspects of welfare: Whenever I am speaking solely of either one of these, I will be explicit about my doing so.

It might help if we now briefly return to Nikhil and the state of affairs *Nikhil's drinking/metabolizing milk*. Suppose that Nikhil is driving home from work at, say, 6:00 pm, and that he doesn't yet have any desire for this state. At this point in time, then, this state is (on DF theories) a far-possible aspect of Nikhil's welfare. Now suppose that at,

say, 6:03 pm Nikhil forms the desire for this state. He is still in his car, though, and so can't access his refrigerator. At this point in time, then, this state is (on DF theories) a merely near-possible aspect of Nikhil's welfare. Finally, suppose that at, say, 6:10 pm Nikhil is at home and is actually drinking/metabolizing milk. At this point in time, then, the state of affairs *Nikhil's drinking/metabolizing milk* is (on DF theories) an actual aspect of Nikhil's welfare. Similar remarks can be made about OL theories. Given that Nikhil's stomach is fairly empty (and perhaps also given some other assumptions), it follows (on OL theories) that the state of affairs *Nikhil's drinking/metabolizing milk* now falls under the basic good of health for Nikhil. Moreover, given that Nikhil is now in his car and so isn't actually drinking/metabolizing milk, it follows (on OL theories) that this state is now a merely near-possible aspect of Nikhil's welfare. Now suppose that it is a few minutes later, that Nikhil is home, and that he is now drinking/metabolizing milk. It follows (on OL theories) that the state *Nikhil's drinking/metabolizing milk* is now an actual aspect of Nikhil's welfare. Finally, suppose that Nikhil has just eaten a full meal, that his stomach is very full, and that drinking/metabolizing milk would be gluttonous. At this point in time, the state *Nikhil's drinking/metabolizing milk* does not (on OL theories) fall under the basic good of health for Nikhil. Thus, at this point in time, the state *Nikhil's drinking/metabolizing milk* is (on OL theories) a far-possible aspect of Nikhil's welfare.

When OL theorists speak of what is prudentially good for A, they sometimes claim that A's welfare is composed of those states of affairs that are, for her, instances of the basic goods. But they often claim something else: They often claim that A's well-being is composed of the basic goods. It isn't hard to see why this happens. OL theories

operate on two levels: the level of the basic goods and the level of the instances of the basic goods. And, when talking about their theories, OL theorists tend to go back and forth between these two levels. This going back and forth isn't a problem, just so long as everybody is clear about what is going on. With respect to the near-possible/far-possible distinction, it is worth stressing that, when OL theorists claim that A's welfare is composed of the basic goods, what they mean is that the basic goods (accomplishment, friendship, health, etc.) are always near-possible components of A's welfare – that is to say, they mean that A's engaging in accomplishment, or friendship, or whatever is always beneficial to A, and that this is so irrespective of whether or not A is currently engaging in accomplishment, or friendship, or whatever.

One last point here: We shouldn't think that OL theories entail that every near-possible aspect of A's welfare is timelessly a near-possible aspect of A's welfare. Here we can consider the case with Nikhil: Although OL theories entail that health is always a near-possible component of Nikhil's welfare, OL theories also entail that something that now falls under health for Nikhil (say, the state *Nikhil's drinking/metabolizing milk*) need not always fall under health for Nikhil (or need not always be a near-possible aspect of Nikhil's welfare). The general point here is this: Although the basic goods are always near-possible components of A's welfare, the states that now fall under them for A need not timelessly be included within A's welfare (as near-possible components of it).

## **5. Desires**

Here (in section 5) I will do three main things. First, I will briefly describe what Timothy Schroeder refers to as the standard theory of desire, which is the theory of desire

that I accept. Second, I will make it clear that DF theorists and OL theorists accept the standard theory of desire or at least something very close to it. Third, I will say something about extrinsic (or instrumental) desires.

In Schroeder's book *Three Faces of Desire* he discusses (and actually rejects) what he refers to as the standard theory of desire.<sup>7</sup> According to Schroeder, the standard theory says that to desire that *P* (where *P* is a proposition) is "to be disposed to bring it about that *P*".<sup>8</sup> Moreover, the standard theory (as described by Schroeder) states: "Desires can play all sorts of causal roles within our heads, but there is just one role definitive of desiring, and that is of engaging the mental machinery in such a way as to tend to bring it about that *P*, for the *P* that is the content of the desire."<sup>9</sup>

There are various questions that we might ask about the standard theory of desire. One of them is this: Why is the theory framed in terms of desires for *propositions* (that is, for propositions to hold)? I take it that, strictly speaking, people don't desire propositions – what they desire is that certain things happen, or that certain states of affairs obtain (for instance, that they own a new car, or that their favorite team wins). However, once we admit that states of affairs are the objects of desires, there is no problem with the claim that desires are for propositions, for any state of affairs can be expressed in propositional form. One might object that many desires are for objects that aren't states of affairs. For instance, one might say that what A desires is a new car, not the state of affairs *that she own a new car*. Notice, though, that it doesn't distort things to say that what A desires is that she own a new car. (In what follows I will often say that people desire friendship, health, etc. But it doesn't distort things to say that what I mean is that people desire that they engage in friendship, health, etc.) In short, then, my main point here is this: It may

seem strange to claim that our desires are for states of affairs (which can be expressed in propositional form), but there isn't any problem with our doing this.

Here is another question to ask of the standard theory: Do we need to feel or to be conscious of our desires? The answer is no. Schroeder comments on this issue: After stressing that the standard theory can accept that we feel and are conscious of some desires, he says that "it is an equally familiar claim made by the standard theory that desires need not appear in consciousness at all, as is the case with most of our standing desires (for the well-being of our loved ones, for example)."<sup>10</sup>

Can the standard theory accommodate the common intuition that beliefs and desires have different directions of fit? The answer is yes. What, though, does it mean to say that beliefs and desires have different directions of fit? The idea here is this. Beliefs have a mind-to-world direction of fit. Our beliefs are supposed to represent the way that the world is, and, when we become aware that our beliefs aren't representing the way that the world is, we should change our beliefs (so as to make them fit the world). Matters are different with our desires. Here the goal is to get the world to change so as to fulfill our desires; the goal, in other words, is to get the world to fit our desires. Thus our desires have a world-to-mind direction of fit.<sup>11</sup> It might be thought that we are speaking in an overly metaphorical way when we talk about the differing directions of fit that beliefs and desires have.<sup>12</sup> But, while it is true that we should be careful about just how far we are willing to take this differing directions of fit talk, it nonetheless should be admitted that this differing directions of fit talk is at least largely right. As Schroeder puts it: "The basic fact described by 'direction of fit' is that beliefs can be true but not fulfilled, while desires can be fulfilled but not true. This is a fact that I do not expect to go away."<sup>13</sup> On

the standard theory desires neither are beliefs nor essentially involve beliefs. Thus on the standard theory there is nothing about a desire per se that can be true or false. The key implication here, then, is that the standard theory has no trouble accounting for the basic fact described by direction of fit talk. This counts in favor of the standard theory, since any adequate theory of desire must capture this basic fact.

I have said that, according to the standard theory of desire, desires neither are beliefs nor essentially involve beliefs. Yet some philosophers claim (1) that to desire X just is to believe that X is (somehow) good or, alternatively, (2) that one's desiring X essentially involves one's believing that X is (somehow) good. The guiding intuition here is that one can't desire X unless she sees – that is, cognitively sees – some sort of value in X. This intuition obviously has some force behind it. But, on reflection, I very much doubt that it is true that one can't desire X unless she believes that X is in some way good. Here two points are in order. First, it seems that there are actual cases where people desire X even though they believe that X is thoroughly bad (or in no way good).<sup>14</sup> For instance, it seems that I sometimes am so mad at people that I believe to have wronged me that I want them not just to be punished for having wronged me, but also to be punished beyond the point of what would be a fair punishment; and in these cases my desire for their being punished more than they deserve is a desire for a state of affairs that I believe to be thoroughly bad. Second, very small babies have plenty of desires, but it is doubtful that they believe that the objects of their desires are (somehow) good. In sum, then, I think that it is best to view desires as being motivational/non-cognitive states – I don't think that we should make the additional claim that desires essentially involve the belief that their objects are (somehow) good.<sup>15</sup>

DF theorists and OL theorists accept the standard theory or at least something very close to it. Consider DF theorists. Some DF theorists – David Lewis is one of them – refer to their DF theories as *dispositional* theories of value.<sup>16</sup> Here the guiding idea is (1) that desires are dispositions to bring about the states of affairs toward which they are directed, and (2) that states of affairs become valuable – or, better put, prudentially valuable – the moment that they are desired. (Lewis’s DF theory is actually much more complicated than these remarks indicate, but let’s not worry about that here. The point here is just that DF theorists hold that desires are motivational states – or are dispositions to bring about the states of affairs toward which they are directed.)

What do DF theorists have to say about the thesis that one cannot desire X unless one believes that X is (somehow) good? Unsurprisingly, DF theorists (at least typically) reject this thesis. Lewis, for example, has argued against this thesis in his articles “Desire as Belief” and “Desire as Belief II”.<sup>17</sup> Those philosophers who claim that A can’t desire X unless A believes that X is (somehow) good ordinarily mean to be claiming that A can’t desire X unless A first cognitively sees that X is (somehow) good – that is to say, these philosophers ordinarily mean to be claiming that A’s belief that X is (somehow) good is what gives rise to A’s desire for X. But this way of thinking about how we form desires – this *desired because first cognitively seen to be valuable* way of thinking – seems antithetical to the DF theory way of thinking about how we form desires. After all, on DF theories our desires make things be prudentially good for us in the first place. Thus, according to DF theories, the correct order is *(prudentially) valuable because desired*.<sup>18</sup> Nevertheless, I should note that there is room for DF theorists to accept the thesis that one cannot desire X unless one believes that X is (somehow) good. The idea

here is that a DF theorist could say: ‘You have to see some sort of goodness in the object in order to form a desire for it. But the goodness that you see can’t be prudential goodness – it has to be some other sort of goodness. Things are prudentially good for you because, and only because, you desire them. But your desire for an object always arises at least in part because you see some sort of non-prudential goodness in it.’ At least typically, though, DF theorists don’t take this position. Indeed, the typical DF theory view is that one can desire X without believing that X is (in any way) valuable.

In light of the foregoing, one might say: ‘The typical DF theory conception of desire is strange. If this view of desire is right, then our desires don’t have the function of tracking (or aiming at) what is prudentially valuable – rather, our desires fix on certain value-free objects and thereby make them be prudentially valuable. But why, if this conception of desire is correct, do our desires fix on certain objects in the first place? The typical DF theory view of desire seems to entail that our desires are blind. In other words, the typical DF theory view of desire seems to entail that our desires fix on objects arbitrarily, or simply by fiat.’<sup>19</sup> Here DF theorists might respond by saying that, since our desires are motivational states (and not cognitive states), we shouldn’t expect them to have the function of fixing on objects for any particular reason – rather, we should expect them to fix on objects spontaneously. In other words, DF theorists might simply grant that it is true that our desires blindly (or spontaneously) fix on objects – but then DF theorists might quickly add that this isn’t a problem, for this is precisely what we should expect from our desires, given the fact that they are motivational/non-cognitive states.

Supposing that our desires do indeed fix on objects in a blind (or spontaneous) manner, it would make sense to think that our desires might fix on some very strange



objects from time to time. With respect to this matter, it may help if we briefly consider something that Warren Quinn says in his article “Putting Rationality in its Place”:

Suppose I am in a strange functional state that disposes me to turn on radios that I see to be turned off. Given the perception that a radio in my vicinity is off, I try, all other things being equal, to get it turned on. Does this state rationalize my choices? Told nothing more than this, one may certainly doubt that it does. But in the case I am imagining, this is all there is to the state. I do not turn the radios on in order to hear music or get news. It is not that I have an inordinate appetite for entertainment or information. Indeed, I do not turn them on in order to *hear* anything. My disposition is, I am supposing, basic rather than instrumental. In this respect it is like the much more familiar basic dispositions to do philosophy or to listen to music.<sup>20</sup>

DF theorists hold that to desire X is to be disposed to make X obtain; and, as it stands, this definition of desire does not rule out one’s being able to desire very strange things – for instance, one’s being able to desire to turn on radios whenever she sees them turned off. Speaking generally, moreover, DF theorists have been willing to admit that it is indeed possible for people to desire very strange things. This issue is an important one for our inquiry, and we will discuss it in some detail in Chapter 1. In short, the main worry for DF theorists here is this. It is hard to believe that something like turning on radios whenever one sees them turned off can be intrinsically *good* for one. However, it appears that DF theorists typically allow that we can desire things such as this, and in turn it appears that DF theorists typically are committed to saying what intuitively seems wrong, namely, that things such as this can indeed be intrinsically *good* for us.

Let’s now turn to the OL theory conception of desire. Mark Murphy, who is an OL theorist, notes that “desires are ascribed to an agent as explaining why that agent acts as he or she does”.<sup>21</sup> Thus Murphy seems to assume, reasonably enough, that desires must at least partly be defined as being motivational states; and this assumption – that is, the assumption that desires essentially have a motivational aspect to them – is one that

OL theorists generally accept without hesitation. The hard question for OL theorists is whether desires essentially involve *not only* a motivational aspect, but *also* a cognitive aspect. Indeed, the hard question for OL theorists is whether one's desiring X essentially involves one's believing that X is (somehow) good. While OL theorists generally hold that *many* desires arise because one believes that their objects are (somehow) good, OL theorists generally hold that it isn't clear that *all* desires arise in this way.

David Brink, who is an OL theorist, seems to capture the view of this matter that OL theorists generally accept when he states:

The desire-satisfaction theory [or DF theory] says that something is valuable just in case, and because, it would contribute to the satisfaction of actual or counterfactual desires. But...this seems to get things just about the opposite way around in many cases; we desire certain sorts of things *because we think these things valuable*. This is true of our preferences for many activities and relations as well as states of the world. It is not that these things are valuable because we desire them; rather, we desire them because we think them valuable...Desire-satisfaction theories get it *just about* the opposite way around in such cases because such desires presuppose only the *belief* that the object of desire is valuable and not the fact that it is valuable. But this fact about the explanation of desire is important, because it shows that *we presuppose* that some things are valuable independently of their being objects of actual or counterfactual desire. I say that desire-satisfaction theories get it the wrong way around in *many* cases, because this sort of value-laden explanation may not be required for all desires. We don't seem to need to posit a belief on Alan's part that baklava is more valuable than other desserts in order to explain his preference for baklava. It seems to be enough that Alan likes baklava more than other desserts.<sup>22</sup>

Does Alan's desire for baklava require that Alan has a belief that baklava is in some way valuable? It seems that Brink is here saying that we don't need to posit any such belief on Alan's part in order for it to be true that Alan desires baklava – this desire *might*, Brink seems to think, just be a brute (or spontaneous) desire.<sup>23</sup>

More generally, Brink seems to hold that, although *many* desires arise from one's having the prior belief that their objects are (somehow) valuable, it isn't clear that *all*

desires arise in this way. And, as I have said, Brink's view seems to be representative of the OL theory camp's view in general: The general OL theory view seems to be that, although *many* desires arise from one's having the prior belief that their objects are (somehow) valuable, it isn't clear that *all* desires arise in this way – thus it is best to remain noncommittal with regard to the thesis that says that one can't desire X unless one believes that X is in some way good. In any case, the main point to take away here is that OL theorists accept the standard theory of desire or at least something very close to it.

What should we say about extrinsic (or instrumental) desires? The standard theory of desire is a theory of what *intrinsic* desires are. Schroeder makes this clear.<sup>24</sup> But what should we say about extrinsic desires? Do we need a theory for them too?

There are philosophers who claim that all desires are intrinsic desires.<sup>25</sup> Murphy, for instance, asks us “to take literally a common manner of speaking: an agent that wants Y solely in order to satisfy his or her desire for X doesn't *really* want Y; what the agent *really* wants is X.”<sup>26</sup> If a dad changes his baby girl's dirty diaper, then it is plausible to think that the dad does this not because he in any way desires to do this, but rather because (1) he desires that his baby girl have a clean bottom and (2) he believes that changing his baby girl's diaper is an effective way of ensuring that his baby girl has a clean bottom. Here, of course, we might *say* that the dad *extrinsically* desires to change his baby girl's diaper – that the dad desires to change his baby girl's diaper not for its own sake, but only as a means. But notice that there doesn't seem to be any *need* to say this: It seems that we can explain what is going on here without appealing to any extrinsic desire that the dad has. I am inclined to agree, then, with those philosophers who hold that there are no extrinsic desires – or, at any rate, that we don't need to speak of them.

Still, I see no need to take any stand on this matter, and so my official position is that I am leaving open the question of whether there are such things as extrinsic desires.

Nevertheless, I will say this. If there are such things as extrinsic desires, it seems that they must differ from intrinsic desires in at least the following respect: Whereas intrinsic desires have motivational force in their own right, extrinsic desires (if there are such things) do not have any motivational force in their own right – that is, extrinsic desires do not by themselves *add* any motivational force.

Here it may help if we consider some comments that Murphy has made. Murphy proposes the following desire-individuation principle: “for all putative desires A and B, A and B are distinct if and only if A and B together would have motivational force in addition to that which either of them alone would have.”<sup>27</sup> And then Murphy adds:

[W]ith regard to the idea of ‘additional motivational force’: what I have in mind, roughly, is that the presence of both desires, rather than one or the other alone, would add to either the *scope* or the *power* of the agent’s motivation, where the scope of motivation is the set of states of affairs toward which the agent is motivated and the power of motivation is the degree to which the agent is motivated to promote some state of affairs. So, with respect to scope: if there is a single state of affairs toward which A and B together motivate yet toward which A alone does not, and a single state of affairs toward which A and B together motivate yet toward which B alone does not, then we will know that we are dealing with distinct desires. With respect to power: if there is a single state of affairs toward which A and B motivate the agent to a greater degree than A alone does, and there is a single state of affairs toward which A and B motivate to a greater degree than B alone does, then we will know that we are dealing with distinct desires. If A and B together would motivate the agent with the same scope and power as A alone would, then B is superfluous, and should not be ascribed to the agent along with A; if A and B together would motivate the agent with the same scope and power as B alone would, then A is superfluous, and should not be ascribed to the agent along with B.<sup>28</sup>

As I have said, if there are such things as extrinsic desires, it seems that they must differ from intrinsic desires in at least the following respect: Whereas intrinsic desires have motivational force in their own right, extrinsic desires (if there are such things) do not

have any motivational force in their own right – that is to say, extrinsic desires do not by themselves *add* any motivational force.

Thus, if we say that the dad in the above example *extrinsically* desires to change his baby girl's diaper, we must mean to be saying that the dad's extrinsic desire to change his baby girl's diaper isn't something that by itself adds any motivational force. All of the motivational force that is present here is flowing to the state *that his baby girl have a clean bottom* – none of the motivational force that is present here is flowing to the state *that the dad changes his baby girl's dirty diaper*. Yes, the dad pursues the state *that he changes his baby girl's dirty diaper*. But still, he isn't motivated by *this* state – he isn't disposed to bring *this* state about. He is motivated solely by the state *that his baby girl have a clean bottom* – indeed, it is just this latter state that he is disposed to bring about.<sup>29</sup>

I think that we have now said enough about extrinsic desires. When it comes to theories of well-being, it is intrinsic desires – not extrinsic desires – that really matter. For instance, when DF theorists say that well-being is a function of (and only of) desires, they are talking about (and only about) *intrinsic* desires. Similarly, when I say that X can't be an aspect of A's welfare unless X's obtaining satisfies a desire that A has, I mean to be talking about an *intrinsic* – and not an extrinsic – desire that A has. (As I indicated earlier, I always use 'desire' to refer to intrinsic desires – unless I note otherwise.)

To sum up, we have done three main things here in section 5. (1) We have briefly described the standard theory of desire, which is the theory of desire that I accept; (2) we have made it clear that DF theorists and OL theorists accept the standard theory of desire or at least something very close to it; and (3) we have said something about extrinsic (or instrumental) desires.

## Chapter One: Desire-fulfillment Theories

### 6. What might be attractive about DF theories?

DF theories assert that X is an aspect of A's welfare if and only if – and directly because – A desires X.<sup>30</sup> Thus DF theorists offer us a *full reduction* of human well-being to desire.

What might be attractive about DF theories? Here it should be said that DF theories might be considered attractive for at least three reasons.

One: The DF theory explanation of human welfare is, we might say, *complete*. It doesn't leave us wondering exactly what the proposal is here. We know exactly what the proposal is here. It is that human welfare has its source in one and only one thing: desire.

Here we might contrast DF theories with OL theories. If the OL theorist claims that A's writing a great novel is an aspect of A's welfare because A's writing a great novel is something that instantiates accomplishment for A, then to an extent we will be satisfied with the OL theorist's explanation. But to an extent we won't be. I say this because we will want to know why accomplishment is on the list of basic goods, and also because we will want to know what it is that unifies the basic goods such that they form one set rather than a disconnected heap. We will discuss this matter further in Chapter 2 (see section 16). But, in short, the main point here is that OL theorists probably need to do more than just offer us a list of basic goods if they want to succeed in providing us with a full account of human welfare. Notice, though, that this problem doesn't arise on DF theories. After all, it is quite clear what the DF theory proposal is. The proposal is that human welfare fully reduces to desire. Thus the notion of desire unifies the DF theory explanation of human welfare in a satisfying way.

Two: DF theories might be thought to be attractive because they fit well with naturalism. Further, it is obvious that many DF theorists – for instance, Lewis and Peter Railton – are strongly motivated, at least in part, by naturalistic concerns.<sup>31</sup> Naturalism in ethics is a large issue, and I won't say much at all about it here. However, I will say that, if one is intent on being a naturalist, then one will likely feel some push away from OL theories and hybrid theories, since, after all, both OL theories and hybrid theories entail that prudential values are not fully reducible to anything non-evaluative. There are, of course, OL theorists who embrace naturalism – Brink is one of them.<sup>32</sup> But the point still holds: If one is intent on being a naturalist, one will likely feel some push away from OL theories and hybrid theories. At that point, then, one will be apt to consider the currently circulating types of welfare theories that fit well with naturalism: welfare hedonism, DF theories, and L.W. Sumner's welfare theory.<sup>33</sup> And, of these three options, the DF theory option is by far the most popular. To sum up: It makes sense for philosophers who are intent on being naturalists – for instance, Lewis and Railton – to become DF theorists.

Three: At least by my own lights, the most attractive feature of DF theories is that they seem to capture the subject-relativity of human well-being (or the *for* part of 'good for') in a fairly straightforward way. As indicated earlier (see section 1): In speaking of someone's welfare, we aren't (or aren't just) speaking of what is intrinsically good – we are speaking of what is intrinsically good *for her*. The phrase 'for her' obviously needs to be captured. In capturing it there are at least two things to keep in mind: individual differences and prudential internalism. DF theories have no trouble explaining individual differences in well-being: DF theories assert that different things are intrinsically good for different people because different people desire different things. With respect to

prudential internalism, the main point is that a welfare theory must provide a sufficiently non-alienating account of human welfare. As indicated earlier (again, see section 1), prudential internalism says that X can't be an aspect of A's welfare unless there is some sort of a positive hookup between X and A. Or, as Railton famously puts it:

[I]t does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him.<sup>34</sup>

The guiding thought behind prudential internalism, then, is that it can't be true that X is *both* intrinsically good for A (or an aspect of A's welfare) *and* thoroughly alienating to A – after all, if X were completely alienating to A, then X couldn't truly be *for* A.

In its most straightforward form prudential internalism says that, if X is an aspect of A's welfare, then A must have some sort of pro-attitude toward X – that is, A must have some favorable mental state (or combination of favorable mental states) vis-à-vis X. Since DF theories entail that all of the aspects of A's welfare are directly constituted by A's desires, DF theories satisfy straightforward prudential internalism. It seems, then, that DF theories provide us with a sufficiently non-alienating account of human welfare.

I don't mean to be suggesting that DF theories don't have any trouble at all with alienation. After all, someone might well be alienated from her own desires in certain cases. For instance, a smoker may desire to smoke, may be mad at herself for having this desire, and may experience a sense of alienation from this desire. Notice, though, that this smoker *is* positively linked to the state of affairs *her smoking* in that she *does* have a pro-attitude (in particular, a desire) toward this state of affairs. *To that extent*, then, she is not alienated from this state of affairs (which, according to DF theories, is a state of



affairs that counts as an aspect of her welfare). The alienation occurs, then, not since and inasmuch as she desires this state of affairs, but rather, say, since and inasmuch as she has con-attitudes that are directed toward her desire for this state of affairs. (Maybe, although she desires to smoke, she desires not to have this desire.) What these remarks bring out, among other things, is the fact that alienation is a complicated phenomenon. And that raises the difficult question of what, exactly, we should expect of a welfare theory when it comes to the attempt to provide us with a sufficiently non-alienating account of human welfare.

This is an important question, and we will be considering it in various places later on (for instance, see Chapter 2, sections 19-21, where I will argue that OL theories can't accept a strong enough form of prudential internalism, and so can't provide us with a sufficiently non-alienating account of human welfare). For now, it is enough if I just note that, all things considered, it does seem that DF theories succeed in providing us with a sufficiently non-alienating account of human welfare. I say this because on DF theories there is always a positive link (in particular, a desire) between an individual and each of the aspects of her welfare. On DF theories, then, there is always a clear sense that can be given to the *for* part of 'good for': The aspect of welfare (whatever it may be) is *for you*, says the DF theorist, since and insofar as you have the positive attitude of desire toward it.

There are many problems that DF theories face. Yet it would not be wise if we were to try to discuss all of them. In the rest of Chapter 1, then, we will focus primarily on just two problems (or, perhaps better put, two types of problems) that DF theories face: the stability problem and the problem of capturing the *good* part of 'good for'.

## 7. DF theories and the stability problem

R. B. Brandt has argued that, because our desires can and do change over time, DF theories are hard to swallow.<sup>35</sup> He states:

Notice that one acquires some desires and loses others as one matures: loses one's desire to be an airline pilot, perhaps, and acquires one to provide for one's family. There are temporary fancies: a person suddenly wants to learn French, works at it, and then loses interest before achieving mastery. Most notably, some desires are cyclical, in the sense that after satisfaction there is a period of no desire for a whole family of events, followed by a recovery of interest... In view of these facts, what is a would-be maximizer of satisfaction of desires to do? ... Does the length of time a person entertains a desire make a difference, so that a sadistic wish for an angry hour counts less than a wish entertained for a whole month (say, only 1/720 as much)? You might say that such comparisons are irrelevant; it is desires at the time of satisfaction that count. But consider an objection to this: a convinced skeptic who has rebelled against a religious background wants, most of his life, no priest to be called when he is about to die. But he weakens on his deathbed, and asks for a priest. Do we maximize his welfare by summoning a priest? Some would say not, in view of his past desires.<sup>36</sup>

When faced with the fact that our desires change over time, DF theorists typically shift to (what we can refer to as) *present moment* DF theories. Present moment DF theories say that the only desires that constitute one's well-being are those desires that one now has.

Given that our desires change over time, moving to a present moment DF theory is presumably the best option that is open to DF theorists. But still, present moment DF theories are problematic. Here consider the above case with the religious skeptic. The general point that Brandt is trying to make here is that there are cases where, given that we are already assuming that DF theories are true, it appears that it is really one's *past* desires – and *not* one's *present* desires – that determine one's *present* welfare.

Take this case, for example. Suppose that a young academic has been working hard on a book for the last two years and that, throughout the whole writing process, she has had the desire to finish the book (that is, to see it all the way through to publication).

However, now suppose that, as she reads comments from her friends and colleagues on the initial draft of her book, she finds that her friends and colleagues have a number of stinging criticisms to offer. Suppose, moreover, that our young academic is crushed by these stinging criticisms, that she temporarily comes to believe that the book is a failure, and, in turn, that she temporarily loses her desire to finish the book. What should we say about this case? Given that we are already assuming that DF theories are true, and given also that in fact the book is by no means a failure (even if it nevertheless is in need of significant revisions), it makes sense to think that, during this period in which our young academic no longer desires to finish her book, it is actually her *past* desires, and *not* her *present* desires, that determine her *present* welfare (or, at any rate, that determine her present welfare as it relates to the matter of her finishing or not finishing her book).

There is another serious problem with present moment DF theories. How, on a present moment DF theory, are we supposed to make sense of the idea of promoting our own and others' well-being across time? What, exactly, are we supposed to be doing in this regard? Suppose that someone is trying to promote her own well-being across time. Should she have the goal of satisfying the objects of her present desires, the objects of the desires that she believes she will have in the future, the objects of those desires that she has held for a long time now, or what? It is hard to understand what the goal is.

The problem here can be put like this. On present moment DF theories one's welfare is fixed by what one presently desires. However, many of one's present desires are directed toward the future. It may seem, then, that the prudentially rational thing to do is to start to work so as to satisfy the present desires that one has that are directed toward the future. Yet, by the time the future comes (that is, by the time the future

becomes the present), the desires that one was working to satisfy may now be irrelevant (that is to say, these desires may now be such that they don't play any role in determining one's welfare, since, after all, one's desires may have changed over time). Naturally, this problem arises not only with respect to the promotion of one's own welfare across time, but also with respect to the promotion of others' welfare across time. You may start to work so as to satisfy, say, your child's desires for certain future states (with the aim of making your child better off). But it is hard to know whether this is really a reasonable thing to do, since, by the time the future arrives, the desires of your child that originally propelled you to do this may no longer play any role in determining your child's welfare.

With considerations such as those just mentioned in mind, Brandt has said that the DF theory project "is not a plausible, or even an intelligible one, when we work it out in detail."<sup>37</sup> It may be an overstatement to say that the DF theory project is *unintelligible*. But, even so, there is something deeply implausible about the DF theory project. Well-being has a certain fixity to it, a certain stability to it. And this stability is what we do and must look toward when we are considering how best to promote our own and others' well-being across time. OL theories have no significant trouble here: The basic goods are fixed components of one's own and others' well-being. So it is to the basic goods that one looks when one wants to know how best to promote one's own and others' well-being across time. DF theories, in contrast, are in deep trouble here.

## **8. DF theories and sufficient broadness**

When it comes to the *good* part of 'good for', there are two general things to keep in mind. First, a welfare theory must be *sufficiently broad*, which is to say that it must

include within the content of an agent's well-being every state of affairs that truly is an aspect of her welfare. Second, a welfare theory must be *sufficiently narrow*, which is to say that it must exclude from the content of an agent's welfare every state of affairs that isn't in fact an aspect of her welfare. In this section I will briefly discuss sufficient broadness. Then, in sections 9 and 10, I will discuss sufficient narrowness.

DF theories have trouble with sufficient broadness. Return to the case of the young academic discussed just above in section 7. There is a period of time during which she has lost the desire to finish her book, and yet during which finishing the book seems to be included within her well-being. Here 'included within her well-being' should be made clearer, for this phrase can mean 'included within her well-being as a *merely near-possible* aspect of it' or, alternatively, 'included within her well-being as an *actual* aspect of it'. What we should say here is this. There is a period of time during which our young academic has the lost the desire to finish her book (and thus isn't in fact finishing her book), and yet during which finishing her book is included within her well-being as a *merely near-possible* aspect of it.

Aside from the case of the young academic, there are various other cases where DF theories appear to be insufficiently broad. Here consider a quote from Sumner:

Having never heard bluegrass, I chance on a band in the park and find that I like it. Having nursed a long-standing suspicion of the Mediterranean, I am persuaded against my better judgment 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.<sup>38</sup>

The bluegrass case can be put like this. Just before A chances on the band in the park, it is true that A's hearing bluegrass is included within A's welfare as a *merely near-possible* aspect of it; and yet A has no desire for this *merely near-possible* aspect of his welfare,

which implies that DF theories have this case wrong. The Mediterranean case can be spelled out similarly. Friends of A persuade him to go to the Mediterranean with them. Although A doesn't desire to go to the Mediterranean, he does go. He goes solely as a means of fulfilling his desire to spend some time with his friends. As it turns out, though, A finds that he loves the Mediterranean. Here, then, we should say: Just before A goes to the Mediterranean, it is true that A's going to the Mediterranean is included within his welfare as a *merely near-possible* aspect of it; and yet A has no desire for this *merely near-possible* aspect of his welfare, which implies that DF theories have this case wrong.

In all three of these cases – the case of the young academic, the bluegrass case, and the Mediterranean case – the root of the problem is the agent's *beliefs*. The young academic loses her desire to finish her book because she falsely believes that the book is a failure. If not for this false belief, she would have had the desire to finish her book. In the bluegrass case the agent doesn't know about bluegrass music. But, had he known about it, he would have had the desire to listen to it. And in the Mediterranean case the agent lacks the desire to take a vacation to the Mediterranean because he has (roughly put) the false belief that the Mediterranean is an unattractive vacation spot.

Before leaving section 8 behind, I should stress that, when it comes to DF theories and sufficient broadness, there is one common type of objection from OL theorists that, to my mind, does not go through. OL theorists often object to DF theories as follows: DF theories entail that, if someone lacks the desire for friendship, or accomplishment, or health, etc., then it follows that this good is not included within her well-being – yet that seems wrong, for surely this good is included within this person's well-being. Here DF theorists may respond as follows. 'It is simply question-begging for OL theorists to say

that DF theories are implausible inasmuch as DF theories entail that, if someone lacks the desire for, say, friendship, then friendship isn't part of her welfare. The right thing to think here is this. It is certainly true, for the vast majority of people, that friendship, accomplishment, etc., are part of their welfare. But that is simply because the vast majority of people desire friendship, accomplishment, etc. With respect to those people who happen to lack the desire for, say, friendship, we should say that friendship isn't part of their welfare. After all, if someone doesn't want to engage in friendship, then why should we think that engaging in friendship benefits her?

Yet this response is unsatisfying, at least to most of us. So as to try to gain some clarity on this matter, it will help, I think, if we draw a distinction between two sorts of cases: the actual/close to actual sort and the far from actual sort. By 'actual/close to actual cases' I intend non-fanciful cases. As for 'far from actual cases', I intend fanciful cases: Yes, we can imagine these cases, but we have no good evidence that they can occur in our world (or, more specifically, our environment) or in any world/environment that is relevantly like ours (say, a world/environment where the same laws of nature are operative, and where the history is substantially similar). It seems that this distinction matters for the debate over theories of human well-being. I say this because the concept of human welfare that we are all theorizing and arguing about (that is, our everyday concept of human welfare) may well apply only to humans in our world/environment and to humans in worlds/environments that are relevantly like ours. Beyond that, the concept may well have no legitimate application. What I propose – and I mean for this proposal to hold not just here, but throughout our entire inquiry – is that we simply bracket all far from actual cases. I propose, that is, that we focus our attention solely on actual/close to

actual cases. (Naturally, I am well aware that some philosophers will not like these restrictions: They will want a welfare theory that covers all imaginable cases. But, while there is, perhaps, nothing wrong with wanting this, we should tone down our aspirations, at least for the time being. It is definitely a tall enough order, at least for now, if we simply try to come up with a welfare theory that can handle all of the actual/close to actual cases.)

In arguing for the no priority theory in Chapter 3, I am going to try to show that there are very few actual/close to actual cases that involve people who lack the desire for one or more of the basic goods (that is, for friendship, or for accomplishment, or for health, etc.). Also, I am going to argue that, with respect to those very rare actual/close to actual cases that involve people who lack the desire for one or more of the basic goods, we don't have good reasons for thinking that the non-desired basic goods are part of the welfare of these people.

Given that this argument of mine is on target, it follows that DF theories are unscathed by objections from OL theorists that ask us to consider people who lack the desire for friendship, or for accomplishment, or for health, etc. Therefore, when it comes to these sorts of objections to DF theories, I do not think that DF theories are in any trouble.

Nevertheless, it does seem that DF theories have at least some trouble when it comes to sufficient broadness. I say this because it seems that DF theories have at least some trouble with cases such as the case of the young academic, the bluegrass case, and the Mediterranean case. As we will see later (see sections 12-14), DF theorists typically try to deal with cases such as these by moving to hypothetical DF theories.



## 9. DF theories and sufficient narrowness: four objections

The most common way of objecting to DF theories is to claim that they are not sufficiently narrow. In other words, the most common objections to DF theories are ones that focus on desires that appear to be aimed at objects that are not intrinsically good for the agents in question. Often these objections are referred to as *defective desire objections*. Contrary to what DF theorists may tell us, it seems that we can and do have defective desires. Here (in section 9) I am going to discuss four defective desire types of objections to DF theories: (1) objections that involve desires for things like money and fame, (2) objections that involve desires that flow from addictions (for instance, the desire to smoke), (3) objections that involve compulsive desires (like the desire to wash one's hands even though they are already clean), and (4) objections that involve bizarre desires (say, the desire to collect plates of mud, or to eat glass). With respect to each of these types of objections, the first question to ask is whether there really are actual/close to actual cases that involve people who have the sorts of desires in question. (As I noted earlier, I always use the term 'desire' to refer to intrinsic desires – unless I explicitly note otherwise; and, as indicated in section 5, it may well be that what we commonly refer to as intrinsic desires are in fact the only desires there are.)

One: Are there actual/close to actual cases that involve people who have desires for things like money and fame? Here we might consider something that Schroeder has said:

[O]ne could object that people do not generally intrinsically desire money. My own experience of average Canadian and U. S. citizens, however, suggests otherwise. Only a few people have *powerful* intrinsic desires for money, no doubt, but it is not for nothing that love of money is said to be the root of all evil.<sup>39</sup>

Here Schroeder isn't using 'desire' in quite the same way that I use it (or in quite the same way that the standard theory of desire uses it). But, supposing that Schroeder were here using 'desire' as I use it, his point would still hold. For, although people in general don't have powerful desires for money (people such as Donald Trump would count as exceptions to this rule), it certainly seems as though people in general do desire money, at least from time to time. Most of us are well aware, of course, that money has no intrinsic worth – that its value is purely instrumental. But that doesn't seem to be enough to stop us from wanting it, at least from time to time. This raises the question: Why, if we know that money has no intrinsic worth, do we desire it?

J. S. Mill's commentary on this matter is justly famous. Regarding the desires for money, power, and fame, Mill says: "What was once desired as an instrument for the attainment of happiness, has come to be desired for its own sake".<sup>40</sup> The main idea here, then, is that *the motivational force* that flows to certain states of affairs that money is a necessary instrument to (for instance, certain states of affairs that instantiate health and pleasure for people) is somehow *transferred* to the having of money. Speaking for myself, I know that I spend a fair amount of time *worrying* about money. This is not at all uncommon for graduate students – or, for that matter, for people in general. Further, in my own case the motivation to have money seems to be purely instrumental some of the time – but, that said, I have little doubt that there are stretches of time where I desire money. What happens during these stretches of time, I think, is that my motivational force that flows to, for instance, the state of affairs *that I can pay rent next month without having to ask my parents for money* somehow transfers to (or overflows to) the state *that I have a substantial amount of money*. Also, I take it that my own case is not uncommon.

Indeed, I take it that the overwhelming majority of people do desire money, at least from time to time.

Two: Are there actual/close to actual cases that involve people who have desires that flow from addictions – say, the desire to smoke? When many people first begin to smoke, it may well be that they don't pursue the act of smoking for its own sake. For instance, it seems that many people start out smoking in high school simply for social reasons – that is, because their friends do it, and it is something that they and their friends can do together. Here, then, it would be the engagement in friendship that one desires, and one pursues the act of smoking solely as a means. Further, even with respect to those people who have been smokers for years on end, it is plausible to think that, at least sometimes, their motivational force flows not to the whole act of smoking – rather, it flows only to the pleasure/relaxation that they derive from smoking. At some point, though, the argument that smokers *never* desire *the whole act* of smoking begins to wear thin. Again here, Mill's comments concerning desire-transferences are key. If Smokey has been a smoker for years, then it seems likely that, at least much of the time, Smokey has the desire for the whole act of smoking; and, further, it seems likely that this desire is there (when it is there) because there has been a certain transference (or overspill) of Smokey's motivational force from, say, the state *Smokey's feeling pleasure/relaxation derived from the act of smoking* to the state *Smokey's engaging in the act of smoking*.

Three: Are there actual/close to actual cases that involve people who have compulsive desires – for instance, the desire to wash one's hands even though they are already clean? The case of compulsions is importantly different from the sorts of cases discussed above. In the above sorts of cases, there usually aren't any cognitive problems

that must be considered. After all, people who desire money usually already know that money has no intrinsic worth, and people who smoke usually already know that smoking is an unhealthful thing to do. Yet, with respect to compulsive people, there are certain cognitive problems – in particular, certain *doubting* problems – that must be considered.

Two quotes from Judith Rapoport – a physician who has worked with compulsive patients for many years – should help here. The first quote:

Some patients are “checkers,” they check lights, doors, locks – ten, twenty, or one hundred times – or repeat peculiar acts over and over again...But most often, the patient is a “washer” who feels he must wash over and over again. All of these problems have common themes: you can’t trust your ordinary good judgment, can’t trust your eyes that see no dirt, or really believe that the door is locked...The compulsion keeps coming back, and you ask yourself “Do I really know? I still feel something is wrong.”<sup>41</sup>

And the second quote:

Obsessives have lost their ability to “know” certain simple things that we all take for granted, things that we all constantly check for by some mechanism of which we are unaware, except when it doesn’t work. This mechanism is not working for these patients at a level that astonishes me, for they are asking “How do you know?” about things that we find ourselves hard put to explain: Is the grass really green? Are my eyes blue?<sup>42</sup>

In light of these two quotes, we might be inclined to make the following claim: Those with compulsions have normal desires – indeed, their problems are simply cognitive.

Does this claim make sense? It certainly does if we are talking about *the initial stages* of a compulsion. Think of the matter this way. Someone doubts that the door to her apartment is really locked, and so she gets up and locks it (say, because she wants to be safe in her apartment). Then, a minute later, she begins to doubt that the door is locked. She thinks to herself: ‘Did I lock the door? Or am I imagining that? Maybe I tried to lock it, but it didn’t really lock. At any rate, I better go check it. It’s not that I really want to check it again. But I better check it again, for I definitely want to be safe

in my apartment.’ In *the initial stages* of a compulsion, then, there is no need to posit the presence of a compulsive desire (say, the desire to engage in door-lock-checking). For we can fully explain what is going on simply by way of appealing to, for example, the desire to be safe in one’s apartment and the recurrent belief that the door might not really be locked. Yet, with respect to someone who has had a door-lock-checking compulsion for months or years on end, the claim that her compulsion is *just* a cognitive problem seems strained. It stands to reason that, for this person, there has been – at some point along the way – a certain transference (or overspill) of her motivational force from, say, the state *that she is safe in her apartment* to the state *that she checks the door lock*. (To be clear, I am not claiming that, over time, the cognitive/doubting problem goes away and that a desire problem takes its place. What I am claiming is that, as the compulsion goes on, the cognitive/doubting problem stays, while a desire problem is *added* to it.)

Four: Are there any actual/close to actual cases that involve people who have bizarre desires – for instance, the desire to collect plates of mud, or to turn radios on whenever one sees them turned off? We know that *curiosity* can give rise to strange desires. For instance, I can remember eating the grass in my yard when I was little. I didn’t desire *that I eat grass*. What I desired was *that I know what it is like to eat grass*. However, it is far from clear that desires like these – bizarre desires that are aimed at the satisfaction of curiosity – are a problem for DF theories. After all, it seems that even OL theorists might well agree that the fulfillment of these sorts of desires at least minimally benefits one (since, after all, it brings one knowledge). What we are looking for here, then, are bizarre desires that are aimed at things that rather clearly don’t seem to benefit one at all.

Here we might try turning to people with psychological disorders. For instance, we might observe a homeless schizophrenic who is burying his belongings – tin cans, a plunger, some old tennis racket that he found, etc. – on the outskirts of town. Or again, we might observe a severely autistic child eating glass on the playground at school. However, these cases are very complicated. The schizophrenic may believe that he is a pirate captain, he may desire to bury/to protect his treasure, and he may believe that he is fulfilling his desire to bury/to protect his treasure (when in fact he is really just burying tin cans, a plunger, an old tennis racket, etc.). But, if that is what is going on, then, while it is clear that we have various defective beliefs on hand, it isn't clear that we have any defective desires on hand. I say that because the desire to bury/to protect one's treasure isn't clearly a defective desire. Or, at any rate, it isn't clearly a defective desire if one already has the belief that one is a pirate captain who possesses a treasure. Further, when it comes to severe autistics and the bizarre behaviors that they sometimes perform, there are debates as to what, exactly, is going on in such cases.<sup>43</sup> My general point here, then, is this: We are looking for actual/close to actual cases that involve bizarre desires that are aimed at things that clearly don't seem to benefit one at all, and looking to people with psychological disorders probably isn't a going to give us what we are after here.

As I noted in Section 5, Quinn asks us to imagine a case where someone desires to turn radios on whenever he sees them turned off. Could this be an actual/close to actual case? It seems that it (or something close to it) could be. But it needs to be filled out in a way that makes it believable. In objecting to DF theories, it isn't fair simply to propose that we imagine certain bizarre cases – for instance, it isn't fair to ask us to imagine that A desires to bang his head against a wall for hours on end, without filling

this in with a believable story. We must have good evidence that the cases that we are imagining can happen in our world/environment or in worlds/environments relevantly like ours. Otherwise it won't be clear that we are imagining actual/close to actual cases.

With respect to Quinn's case, let's suppose that Jarrell is little (say, six or seven years old) and that his parents often fight. This is obviously unpleasant for Jarrell, and so he often turns on the radio in his bedroom so as to drown out the back-and-forth yelling of his parents. At first, Jarrell's motivation to turn the radio on in his bedroom is merely instrumental – he does this just to drown out the fighting. But, as time goes by and the habit of flipping the radio on in his bedroom becomes more and more entrenched in Jarrell's motivational structure, some of Jarrell's motivational force *transfers* from the state *that he doesn't hear his parents' fighting* to the state *that he turns radios on when he sees them turned off*. This leads to some strange behaviors from Jarrell. For instance, when Jarrell is in middle school and high school, he spends a good deal of time at his friend Nadir's house. Whenever Jarrell walks into Nadir's bedroom, Jarrell flips the radio on right away. Nadir notices that Jarrell almost never actually *listens* to the radio, and so Nadir sometimes asks Jarrell why he always goes out of his way to flip the radio on. Jarrell always answers: 'I don't know why. It's just something I do.' It seems that the best way to make sense of what is going on here is just to say that Jarrell has a desire to flip radios on whenever he sees them turned off. (Or, alternatively, we might say that the desire here is a desire to flip *bedroom* radios on whenever he sees them turned off – but let's leave this complication to the side.) This desire is bizarre, no doubt – indeed, it is a weird quirk of Jarrell's motivational structure. But, that said, Jarrell's having this desire is understandable, given what we know about Jarrell's history.

So far here (in section 9) I have been trying to make it clear that people in our world/environment can desire money, can have desires that flow from addictions, can have compulsive desires, and can have bizarre desires. But we might now ask whether DF theorists will agree with me about all of this. The answer is that they almost certainly will. Here is why I say this. As indicated in section 5, DF theorists generally reject the view of desire that says that we first cognitively see that things are valuable and then, after that and because of that, form desires for them. Indeed, the typical DF theory view is that the correct order is not *desired because cognitively seen to be valuable* – rather, the correct order is (*prudentially*) *valuable because desired*. Railton notes that “our first-order desires may press upon us willy-nilly”.<sup>44</sup> And this is an apt description of the view of desire that DF theorists generally work with. I say that because DF theorists just do generally assume that our desires (or, at any rate, our first-order, actual desires) are willy-nilly occurrences – that is to say, DF theorists just do generally assume that our desires spontaneously (or blindly) fix on certain value-free objects, thereby making them be prudentially valuable. Given, then, that this is the view of desire that DF theorists generally work with, it would not make much sense for DF theorists to protest the claim that it is possible for the desires of people in our world/environment to fix on objects such as the having of money, or the act of smoking, or the act of checking the door lock yet again, or the act of turning on radios that are seen to be turned off.

We should now ask what any of this shows. To my mind, this shows that DF theories cannot adequately capture the *good* part of ‘good for’. Start with the money case. According to DF theories, if A desires to have money, then A’s having money adds to A’s well-being. But this is hard to swallow. For, while A’s having money may well



be instrumentally good for A, it is hard to believe that A's having money *in itself* makes A better off. Or again, suppose that A desires to smoke. It is very hard to believe that smoking is intrinsically *good* for A. Indeed, the truth seems to be that smoking is (in itself) unhealthy for A and thus (in itself) bad for A. Remember: We aren't here talking about A's deriving pleasure/relaxation from the act of smoking. We are talking about *the act* of smoking. The same holds for the desire to check the door lock yet again, and also for the desire to turn on radios whenever one sees them turned off. Checking the door lock yet again is not intrinsically *good* for one; and, likewise, turning radios on whenever one sees them turned off is not intrinsically *good* for one. The problem, in all of these cases, is that there doesn't seem to be any *value* of any sort in the desired object.

What might DF theorists say in response? They might say: 'The above objections assume that there are some fairly strong constraints on prudential values. Yet why should we think that prudential values are subject to any fairly strong constraints? No doubt there are some fairly strong constraints on, say, moral values. After all, not just anything – not just any action, disposition, policy, etc. – can be morally valuable. However, when it comes to prudential values, matters seem very different. Indeed, it seems that anything can be prudentially good for someone – she just has to desire it to make it be so.'

This response is less than convincing. When we make judgments about, or ask questions about, what is prudentially good for others or for ourselves, we typically assume that there are some fairly strong constraints in play – or, at any rate, we typically assume that A's desiring X does not by itself make X be intrinsically good for A. Of course, it is possible (for all I know) that we are mistaken in typically assuming that A's desiring X does not by itself make X be intrinsically good for A. But I doubt that we are.

Here it seems that the force of the relevant counter-examples is overwhelming. If Jarrell desires to turn on radios whenever he sees them turned off, then it follows on DF theories that Jarrell's turning radios on whenever he sees them turned off is intrinsically *good* for him. Yet that doesn't make sense. Yes, we can see that Jarrell's doing this is *for* him (in that he desires it). But we can't see that his doing this is intrinsically *good* for him. Naturally, there may be value for Jarrell in not having to hear his parents fight. But that is different. Here the question is whether there is value for Jarrell in the bare act of flipping radios on whenever he sees them turned off. And it seems clear that there is no value for Jarrell in this act.

DF theorists may say that all of this is question-begging – that these counter-examples are convincing only to those who have already rejected DF theories. But I disagree. The point of the counter-examples is to get us to start thinking hard about what does and what doesn't (and what can and what can't) fall under our everyday concept of human welfare. If someone desires to turn on radios whenever he sees them turned off, then it follows on DF theories that his doing this is intrinsically *good* for him. But, when we start thinking hard about how well this squares with our everyday concept of human welfare, we can see that it doesn't square with it well at all – indeed, there is a terrible mismatch here. And one need not have already rejected DF theories in order to see that there is a terrible mismatch here. After all, one might be inclined to accept DF theories and then, after considering this case, come to realize that DF theories are implausible.

One final point here: I can see a DF theorist saying that there is nothing obviously implausible about the claim that, if A desires money, or smoking, or door-lock checking, or radio-turning-on, etc., then this thing (whatever it happens to be) is, in some minimal

way, prudentially good for A. To my mind, this makes no sense, since (barring strange circumstances) I see no intrinsic value of any sort in things like money, smoking, door-lock checking, and radio-turning-on. But, just for the sake of argument, let's grant that these sorts of things can be, in some minimal way, prudentially good for people. There is still a major problem here for DF theories. DF theorists will surely hold that the degree to which X is prudentially good for A is a function of the strength with which A desires X. And there are a fair number of actual cases where people strongly desire money, or smoking, or door-locking checking, etc. In these cases, then, DF theorists will end up claiming that money, or smoking, or door-locking checking, etc., *very much* benefits the agent in question (whoever she may be). And that, I think, is terribly implausible.<sup>45</sup>

#### **10. DF theories and sufficient narrowness: a fifth objection**

We should now consider one more type of objection that says (in effect) that DF theories are not sufficiently narrow. This type of objection is perhaps the most common type of objection to DF theories. James Griffin is an OL theorist, and, in objecting to DF theories, he says: "But what relevance have peoples' *actual* desires to what is in their *interest*? One of the discouraging facts of life is that one can get what one actually wants only to find that one is not better off, and sometimes even worse off."<sup>46</sup> Here Griffin appears to have in mind certain everyday cases that seem to show that we actually desire things that are not intrinsically good for us. Dennis might desire a certain job and then get it, only to find that he hates it. Or again, Susie might desire to be in a romantic relationship with Shane and then have her desire fulfilled, only to find that being in this relationship with Shane is on the whole very bad for her. Let's refer to these types of

objections to DF theories as *Griffin-style objections*. There are different ways in which DF theorists might try to deal with Griffin-style objections. I will discuss three such ways.

One: Consider the following case. ‘Dennis desires to work at such-and-such a law firm and then gets the job, only to find that he hates it. Here, then, Dennis’s desire is fulfilled, but his well-being is not advanced.’ With respect to this case, DF theorists might say: ‘The case has been misdescribed. In truth the case runs as follows. There are pros and cons to Dennis’s job. The pros are the interesting cases that he sometimes works on, those few pleasant co-workers of his, etc. And the cons are the long hours, the boring cases that he sometimes works on, his super-cranky boss, etc. Dennis likes the pros of his job and dislikes the cons of his job. Also, we know that Dennis *on the whole* hates his job, and presumably this is so because the various aspects of his job that he dislikes somehow greatly outweigh the various aspects of his job that he likes. Does it make much sense to think that (prior to getting the job) Dennis ever had a desire for any of the various aspects of his job that he now dislikes? No, it doesn’t. At least on that score, then, there isn’t any problem with Dennis’s desires. Further, it makes sense to think that Dennis never desired the state of affairs *his working at this law firm* (full stop). Rather, what he desired (prior to getting the job) were just certain *aspects* of the job (say, to work on interesting cases and to work with pleasant people). Dennis was motivated to take the job, then, not because he desired to take the job (full stop), but rather because he desired certain aspects of the job and because he knew that he couldn’t get these aspects without taking the job. The bottom line here, then, is that this case doesn’t involve any desires that, when fulfilled, fail to advance Dennis’s well-being.’

The above response is meant not only to cover this case with Dennis, but indeed to cover *all* cases where people are motivated to pursue jobs, romantic relationships, etc. But it is precisely the intended generality of this response that makes it implausible. Yes, there are people who are motivated to pursue certain jobs simply because they desire certain *aspects* of these jobs. And the same holds true for the pursuit of certain romantic relationships. It isn't admirable, but many people pursue certain romantic relationships not because they love the other person as such and desire to be with the other person (full stop), but rather just because they desire, say, the financial security that the other person has to offer, or because they desire the other person sexually, or whatever. However, it goes too far to say that people *never* desire to have certain jobs (full stop), that people *never* desire to be in certain romantic relationships (full stop), etc. Moreover, the truth seems to be that it often happens that people desire to have certain jobs (full stop) or to be in certain romantic relationships (full stop), that people have their desire fulfilled, and that they then find that they are on the whole worse off for having their desire fulfilled.

There is another point to mention here. Even if there were some plausibility to the claim that people only desire certain *aspects* of jobs, romantic relationships, etc., this is not the sort of claim that a DF theorist is apt to make. As I noted in section 9, DF theorists typically hold that our desires spontaneously (or blindly) fix on certain value-free objects and thereby make them be prudentially valuable. But, given this view of how our desires operate, it doesn't make sense for DF theorists to advance the response that I outlined just above. The response that I outlined just above says: 'People never really desire things like having such-and-such a job or being in such-and-such a relationship. All that people ever really desire are certain *aspects* – indeed, certain

*positive* aspects – of having such-and-such a job and of being in such-and-such a relationship.’ According to this response, then, our desires always zero in on the positive, and only the positive, aspects of the experiences that are open to us. Yet, if this is the right view of desires, then our desires are anything but the spontaneous (or blind) occurrences that DF theorists typically deem them to be. To sum up, then, the response from above isn’t one that any typical DF theorist is likely to make, for the response from above very much cuts against the typical DF theory view of desire.

Two: With respect to Griffin-style objections, there is another move that DF theorists might make. They might say this: ‘Go back to Dennis. By working at the law firm Dennis fulfills some desires – for instance, the desire to meet some nice co-workers, and also the desire simply to work at the law firm. Inasmuch as these desires are fulfilled, Dennis’s welfare is advanced. However, by working at the law firm many of Dennis’s other desires are frustrated – for instance, the desire to have good hours, the desire to work on certain sorts of law cases, and the desire to have a nice boss. Further, once all of the relevant desire-satisfactions and desire-frustrations are taken into account, it will be clear to us that, as far as this case goes, Dennis’s total desire-frustration outweighs Dennis’s total desire-satisfaction. Or, at any rate, this will be clear to us at least if we make certain assumptions about the relative strengths of the desires that Dennis has in this case. In any case, the bottom line here is this. Inasmuch as Dennis’s desire to work at the law firm is fulfilled, his welfare is advanced. And the net loss in welfare that Dennis suffers as a result of working at the firm can be explained by DF theories without any problem, for this net loss can be explained in terms of total desire-frustration outweighing total desire-satisfaction.’

In advancing this response DF theorists have gone ahead and helped themselves to certain assumptions concerning the relative strengths of the desires that Dennis has in this case. And it isn't clear that it is legitimate for DF theorists to do this. But, even if we leave this matter to the side, there is still a major problem with this response from DF theorists. This response asserts that Dennis's working at the law firm is something that (in itself) adds to Dennis's well-being. Yet, unless one already has a prior commitment to the DF theory, there is no good reason to think that this assertion is true. After all, Dennis is miserable at the firm. Dennis and his job are not at all suited (or fit) for each other. So, unless one is already a DF theorist, one will inevitably conclude that working at the firm is *not* (in itself) adding to Dennis's well-being. DF theorists must convince the rest of us that this natural conclusion is mistaken. However, there really is nothing constructive that DF theorists can say to the rest of us here. Naturally, DF theorists might stomp their feet and say that it is simply a fact that all instances of desire-fulfillment add to welfare. But that won't be convincing to anybody who isn't already a DF theorist.

Three: I know of one more way in which DF theorists might try to answer Griffin-style objections. In his article "The Problem of Defective Desires" Chris Heathwood has argued that his DF theory can handle Griffin-style objections. Heathwood says:

Importantly, I mean the theory above to require *concurrency*: in order for a state of affairs to count as a genuine instance of desire satisfaction, the state of affairs desired must obtain *at the same time* that it is desired to obtain. If I desire fame today but get it tomorrow, when I no longer want it, my desire for fame was not satisfied. A desire of mine is satisfied only if I get the thing while I still desire it, and continue to have the desire while I'm still getting it.<sup>47</sup>

With respect to the case involving Dennis, Heathwood might say: 'DF theorists need not worry about the case with Dennis, since it does not satisfy the *concurrency* requirement.

After all, the state of affairs *that Dennis works at the law firm* does not obtain at the same

time that Dennis has the desire for this state of affairs. Yes, *before* Dennis works at the law firm, Dennis has the desire for this state. However, by the time Dennis works at the law firm, he no longer has the desire for this state (since, after all, he hates his job).’

Heathwood is very forthcoming about his reasons for building concurrence into his DF theory.<sup>48</sup> He says:

Oscar Wilde quipped that the only thing worse than not getting what one wants is getting what one wants. Sidgwick lamented a related point, noting that the objects of our strongest desires often come to us as ‘Dead Sea apples’, no longer wanted once they are gotten, ‘mere dust and ashes in the eating’. Sidgwick thought that moving to an ideal desire theory, one requiring full information about the objects of desire, could solve the problem, since we could learn in advance whether the objects of our desires would come to us flat...But actualist desire-satisfaction theories [such as Heathwood’s DF theory] can accommodate the Dead-Sea-apple phenomenon. They can do so by building concurrence into the theory, as I have done above. 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.<sup>49</sup>

There are at least three problems with Heathwood’s way of dealing with Dead Sea apple cases – or, in other words, Griffin-style cases. These three problems are the merely near-possible problem, the overlap problem, and the retrospective Dead Sea apple problem.

*The merely near-possible problem:* When Sidgwick, Griffin, and others advance these Dead Sea apple cases, they are thinking (1) that A’s desires can and do sometimes aim at things that are not in fact intrinsically good for A, and (2) that we can see that this is so when we reflect on the fact that A sometimes gets what she wants, but then finds that this thing is disappointing to her (or is not what she anticipated it to be). Heathwood answers: ‘But, since A doesn’t want the thing once she gets it, the DF theory is in the clear here – or, at any rate, the DF theory is in the clear if we build *concurrence* into the DF theory.’ However, it seems that the claim that Sidgwick, Griffin, etc., are making still



stands: Our desires can and do go wrong (that is, they can and do aim at things that aren't intrinsically good for us), and these Dead Sea apple cases nicely bring this out. With respect to these Dead Sea apple cases, we should say: Even if A's desire for X (which is not intrinsically good for A) disappears the moment that A actually gets X, this does not change the fact that A's desires *were* going wrong (or *were* aiming at something – namely, X – that was not in fact intrinsically good for her) *just before* she got X.

So as to be more precise about what is happening in these Dead Sea apple cases, we can think in terms of *merely near-possible* aspects of welfare. Consider the case with Dennis. Suppose that at t1 Dennis forms the desire for the state *that he works at the firm*. It seems to follow on DF theories that this state is a merely near-possible aspect of Dennis's welfare at t1. But that is hard to swallow. In order for this state really to be a merely near-possible aspect of Dennis's welfare at t1, this state must be prudentially valuable for Dennis at t1. But why should we think that this state really is prudentially valuable for Dennis at t1? We know that, as things turn out, Dennis works at the firm from t2 indefinitely onwards; and we also know that, as things turn out, Dennis is miserable while he works at the firm. In other words, then, we know that, when this state does obtain at t2, it is *not* prudentially valuable for Dennis. But, if we know that this state is *not* prudentially valuable for Dennis when it obtains at t2, then why should we think that this state *is* prudentially valuable for Dennis at t1 (that is, just prior to its obtaining)? There really is no good reason to think that this state is prudentially valuable for Dennis at t1. If I am right about this, then it seems that DF theories have this case wrong. After all, it seems that DF theories entail that this state is prudentially valuable for Dennis at t1 (or is a merely near-possible aspect of Dennis's welfare at t1).

I am aware of two responses that Heathwood (and other DF theorists) might offer here. Let's refer to these two responses as response #1 and response #2, respectively.

Response #1: Heathwood could try to say (1) that it is legitimate for DF theorists to deny that there are such things as merely near-possible aspects of welfare, and (2) that it follows from this that DF theorists need not worry about the merely near-possible problem. But why should we think that this denial would be legitimate? The root idea of the DF theory is that our desires, and only our desires, make states of affairs be aspects of our welfare. Further, we can and often do desire states of affairs that do not obtain at the moment that we desire them. The most natural thing for a DF theorist to say, then, is that these states – these states that are desired but do not obtain at the moment that they are desired – are prudentially valuable for the individuals who desire them (or are merely near-possible aspects of welfare for the individuals who desire them). To say anything else would be out of keeping with the general tenor of the DF theory. Thus any DF theorist who denies that there are merely near-possible aspects of welfare owes us an explanation as to why his doing this is legitimate (rather than ad hoc).

Still, Heathwood may say: 'You haven't convinced me that I must countenance merely near-possible aspects of welfare. Why, really, must I countenance such things? For something to be a merely near-possible aspect of welfare is for it to be a state of affairs that doesn't now obtain and yet that *is* prudentially valuable for some individual. That is very strange: How can some state that doesn't now obtain nonetheless count as being prudentially valuable for someone?'

To this I say: Suppose that someone is now seventeen – call her Mary. Suppose that Mary now desires, say, to be an architect by the age of twenty-six, and suppose that

Mary does not now desire, say, to be a veterinarian by the age of twenty-six. Given that we are already accepting a DF theory framework, it seems undeniable that Mary's prudential relationship with the state *her being an architect at age twenty-six* importantly differs from Mary's prudential relationship with the state *her being a veterinarian at age twenty-six*. How, then, are DF theorists going to account for this difference? They must find a way to account for this difference: If they simply ignore this difference, then there will be a patent inadequacy in their account. The most obvious way of accounting for this difference is to say that, right now, it so happens that the state *Mary's being an architect at age twenty-six* is a merely near-possible aspect of Mary's welfare, while (in contrast) the state *Mary's being a veterinarian at age twenty-six* is a far-possible aspect of Mary's welfare. Heathwood seems to have two choices here. He can accept that there are such things as merely near-possible aspects of welfare, in which case he can effectively explain what is going on with Mary – or, alternatively, he can refuse to countenance merely near-possible aspects of welfare, in which case he is going to have serious difficulties explaining what is going on with Mary. The upshot of all of this is that it seems that Heathwood can't escape what I am referring to as the merely near-possible problem – or, alternatively, he can escape the merely near-possible problem, but only at the cost of not being able adequately to explain what is going on in cases such as the case of Mary.

Response #2: Still, there is another way in which Heathwood might respond to the merely near-possible problem. According to this second response, Heathwood accepts that there are such things as merely near-possible aspects of welfare, but he takes issue with the way in which I have applied the far-possible versus near-possible distinction to

DF theories. Earlier (see section 4) I noted that ‘X is a near-possible aspect of A’s welfare’ should be taken to mean ‘X is now prudentially good for A’. And I noted that the near-possible aspects of one’s welfare fall into two types: the merely near-possible and the actual types. ‘X is a merely near-possible aspect of A’s welfare’ should be taken to mean ‘X is now prudentially good for A, but X doesn’t now obtain’. ‘X is an actual aspect of A’s welfare’ should be taken to mean ‘X is now prudentially good for A, and X now obtains’. As for far-possible aspects of welfare, ‘X is a far-possible aspect of A’s welfare’ should be taken to mean ‘X isn’t now prudentially good for A, but X might later come to be prudentially good for A.’ Heathwood might agree with all of this and yet disagree with the way in which I have applied these distinctions to DF theories.

Assuming the DF theory structure, my view is (1) that near-possible aspects of welfare are those states that one now desires, (2) that merely near-possible aspects of welfare are those states that one now desires and yet that don’t now obtain, (3) that actual aspects of welfare are those states that one now desires and that now obtain, and (4) that far-possible aspects of welfare are those states that one doesn’t now desire but that one might desire at some point in the future. But Heathwood may say: ‘Your account of how the far-possible versus near-possible distinction applies to DF theories is off target. What we should say is (1) that X is a near-possible aspect of A’s welfare if and only if A would desire X if X were now obtaining, (2) that X is a merely near-possible aspect of A’s welfare if and only if A would desire X if X were now obtaining and X isn’t now obtaining, (3) that X is an actual aspect of A’s welfare if and only if A would desire X if X were now obtaining and X is now obtaining, and (4) that X is a far-possible aspect of A’s welfare if and only if A wouldn’t – but could – desire X if X were now obtaining.’

With regard to Dennis, then, the idea here is (1) that *his working at the firm* is a near-possible aspect of his welfare if and only if he would desire to work at the firm if he were actually working there, and (2) that *his working at the firm* is a far-possible aspect of his welfare if and only if he wouldn't – but could – desire to work at the firm if he were actually working there. We know that, as things turn out, Dennis does not desire to work at the firm once he is actually working there. So it is sensible to think that, prior to his working there, it *isn't* true that his working there is a near-possible aspect of his welfare – indeed, we should think that, prior to his working at the firm, his working there is a far-possible aspect of his welfare. And, with all of that established, it makes sense to think that the so-called merely near-possible problem isn't really a problem for the DF theory.'

The above (possible) line of response from Heathwood attempts to skirt the merely near-possible problem by offering a certain interpretation – for convenience, we'll call it the Heathwood interpretation – of how the near-possible versus far-possible distinction should be applied to DF theories. But the Heathwood interpretation is not without its costs. Keeping in mind that Heathwood is an actual DF theorist (as opposed to being a hypothetical DF theorist), it is worth emphasizing that the Heathwood interpretation severs the link between merely near-possible aspects of welfare and *actual* desires. Indeed, according to the Heathwood interpretation, the actual DF theory entails that X is a merely near-possible aspect of A's welfare if and only if A would desire X if X were now obtaining and X isn't now obtaining. This departs from what we would expect. We would expect the actual DF theory to entail that X is a merely near-possible aspect of A's welfare if and only if A actually desires X and X doesn't now obtain. Also, this departure from what we would expect is problematic, and for at least two reasons.

First, there are plenty of cases where, given that X isn't now obtaining, it seems there is no determinate fact as to whether A would desire X if X were now obtaining. Take Dennis. Whether he would desire to work at the firm if he were actually working there depends on what things would be like for him if he were actually working there. But, given that he isn't now working at the firm, it seems that what things would be like for him if he were actually working there is subject to some considerable indeterminacy. Depending on the types of cases he would be working on, the number of hours he would be working, the clients he would be working with, etc., things could be quite nice for him if he were actually working there (in which case he would have the desire to work there) – or, alternatively, things could be anything but nice for him if he were actually working there (in which case he wouldn't have the desire to work there). All of these things are subject to flux such that it seems that there just is no set fact, prior to his actually working there, as to whether he would desire to work there if he were actually working there. (I will discuss this same general issue in more detail later – see the end of section 14.)

Second, perhaps the most notable virtue of the actual DF theory (as opposed to the hypothetical DF theory, and also as opposed to OL theories) is that it always gives a clear sense to the *for* part of 'good for' – the actual DF theory does this by directly linking all of the aspects of A's welfare to A's actual motivational structure/desire-set. However, at least with respect to merely near-possible aspects of welfare, this notable virtue is lost – or, at any rate, is no longer clearly present – if actual DF theorists accept the Heathwood interpretation. We can elaborate on this matter as follows. If actual DF theorists accept the Heathwood interpretation, then, when dealing with some (any) merely near-possible aspect of A's welfare, they must say to A: 'This state of affairs

(whatever it may be) is *for you* in that it is something that you would desire if it were now obtaining – that said, its being *for you* doesn't depend on there being any actual positive link between it and your actual motivational structure/desire-set.' But it is much clearer and more direct for DF theorists simply to say: 'This state (whatever it may be) is *for you* since and insofar as you actually have the positive attitude of desire toward it.'<sup>50</sup>

*The overlap problem:* With respect to some of these Dead Sea apple cases, it seems that there will be a stretch of time – even if only a short one – where the agent in question is still desiring X, is in fact getting X, and nevertheless is *not* benefiting from X. For instance, when Dennis first starts his job at the firm, it may happen (1) that he still has the desire to work at the firm, (2) that his desire is being fulfilled (since he is in fact working at the firm), and (3) that he isn't benefiting from working at the firm.

This point can be brought out more clearly if we consider that Dennis, looking back on his time spent working at the firm, might say: 'I worked at the firm for a month. The whole experience was terrible. I was miserable there. Did I want to work there? Before I got there, yes, I did want to work there. Also, I would say that I still wanted to work there during my first week on the job. After that first week, though, I lost my desire to work there. I stayed the rest of the month just so as to make sure that I had an income until I found a new job. Why did I still want to work at the firm during my first week there? That is a good question, for I was indeed having a terrible time working there, even during that first week. I initially formed the desire to work at the firm because I believed, wrongly, that this job was an attractive job. Further, I continued to hold this false belief even during my first week of working at the firm. Indeed, it wasn't until after this first week that it finally hit me that this job was in fact an unattractive job. Things

just didn't crystallize in my mind until after that first week on the job. Thus it wasn't until after this first week that I lost my desire to work at the firm. Anyway, it is clear to me now, looking back, that I never benefited from working at the firm – not even during that first week when I still wanted to work there (or, as you philosophers might say, when I still had some of my motivational force flowing to the state *that I work at the firm*).'

The general point that I am trying to make here is this: Although it is doubtful that *all* Dead Sea apple cases involve an overlap problem, it stands to reason that at least *some* Dead Sea apple cases involve an overlap problem. And by 'overlap problem', of course, I mean 'a stretch of time where the agent continues to want X even as he is getting X and is *not* benefiting from it'.

*The retrospective Dead Sea apple problem:* In thinking about these Dead Sea apple cases, Sidgwick, Griffin, and others are thinking about cases where you desire X and then get X, but are immediately, or at least fairly quickly, disappointed by X. There are, however, what we might refer to as retrospective Dead Sea apple cases. In these cases the following happens: The agent desires X, gets X, and continues to desire X and have X for some time – eventually, however, the agent loses her desire for X *and* looks back on her life and makes the judgment that having X was not beneficial to her. Here is an example. Sometimes people who let the pursuit of money dominate certain stretches of their lives later have a conversion of sorts and (roughly) think to themselves: 'Back then, I was deeply consumed by the whole money-making enterprise. I strongly desired to make a good deal of money, and I was in fact making a good deal of money. Also, if you would have asked me back then whether living this life-style was (in itself) good for me, I would have said yes. Now, though, I can see that I was wrong: Making a good deal



of money – living that sort of life-style – was never beneficial to me, not intrinsically anyway.’ Heathwood and other DF theorists must say that the person in this case is just plain wrong when, in looking back on his life, he makes the judgment that he never intrinsically benefited from making a good deal of money (or from living that sort of life-style). But that, I think, is hard to swallow, for it would seem that this person is right.

To sum up: These Griffin-style objections (or Dead Sea apple objections) are a major thorn in the side of DF theorists. At least typically, DF theorists try to deal with these sorts of objections by moving to a hypothetical DF theory – or, roughly, to a theory that says that someone’s well-being is constituted by (and only by) the desires that she would have if she had just been fully and vividly informed with respect to value-free facts. The guiding thought here is that, once agents are fully and vividly informed in this way, their defective desire problems will (at least largely) go away. We will discuss hypothetical DF theories in sections 12-15. First, though, we will discuss second-order DF theories.

## **11. Second-order DF theories**

What motivates the move from a first-order to a second-order DF theory? Here we might turn to some things that Railton and Lewis, respectively, have said about the putative virtues of second-order desires. (So as not to mislead, I should note that Railton and Lewis are both second-order *hypothetical* DF theorists – see section 12.)

Railton thinks, understandably enough, that second-order desires have a critical edge to them that first-order desires lack. In explaining why he rejects the first-order DF theory Railton says: “Yet this theory [the first-order DF theory] is deeply unsatisfactory,

since it seems incapable of capturing important elements of the critical and self-critical character of value judgments.”<sup>51</sup> Moreover, Railton says that “one *embraces* a desire, or accepts it as *goal setting*, when one desires that it be effective in regulating one’s life”, and Railton goes on to say that, although “our first-order desires may press upon us will-nilly”, those “higher-order desires of the sort that are involved in embracing a desire are more responsive to changes in belief” and so “are more closely tied to our identity”.<sup>52</sup>

Lewis holds that, roughly speaking, “values are what we are disposed to value.”<sup>53</sup> This raises the question of what valuing is. Lewis answers that valuing “is some sort of a mental state, directed toward that which is valued.”<sup>54</sup> And then, after rejecting the view that valuing is either a feeling or a belief, Lewis says:

So we turn to desires. But we’d better not say that valuing something is just the same as desiring it. That may do for some of us: those who manage, by strength of will or good luck, to desire exactly as they desire to desire. But not all of us are so fortunate. The thoughtful addict may desire his euphoric haze, but not value it. Even apart from all the costs and risks, he may hate himself for desiring something he values not at all. It is a desire he very much wants to be rid of. He desires his high, but he does not desire to desire it, and in fact he desires not to desire it. He does not desire an unaltered, mundane state of consciousness, but he does desire to desire it. We conclude that he does not value what he desires, but rather he values what he desires to desire.<sup>55</sup>

Like Railton, then, Lewis thinks that second-order desires have a critical edge to them that first-order desires lack.

Speaking generally, it seems that second-order DF theorists hold the following view. ‘In order to attain adequacy, a welfare theory must have some sort of a critical edge to it. This implies that one must be able to check the objects of her current desires against some prudentially authoritative standard – that is, so as to see whether the objects of her current desires really are aspects of her well-being. In determining what sort of standard might be prudentially authoritative for someone (anyone), we need not look to

any standard that is independent of the agent herself. After all, we can claim that, when one wants to know whether the objects of her desires really are aspects of her welfare, she should just ask herself whether she desires to desire what she now desires.’

In assessing whether second-order DF theories are convincing, there are various things we could discuss. I will now briefly discuss sufficient broadness and sufficient narrowness, as well as some problems that are peculiar to second-order DF theories.

*Broadness:* Moving to a second-order DF theory won’t help DF theorists when it comes to broadness. Go back to the three cases we discussed earlier. (1) The young academic loses her desire to finish her book because she believes, wrongly, that the book is a failure. It would not make sense to think that, although she doesn’t desire to finish the book, she does desire to desire to finish the book. (2) In the bluegrass case the agent lacks the desire to hear bluegrass music because he doesn’t know about bluegrass music. But, if he doesn’t know about bluegrass music, then he is not going to desire to desire to hear bluegrass music. (3) In the Mediterranean case the agent believes, wrongly, that the Mediterranean is an unattractive vacation spot. This defective belief leads the agent not to want to vacation in the Mediterranean. It makes no sense to think that, although this agent doesn’t want to vacation in the Mediterranean, he nonetheless does want to want to vacation in the Mediterranean. In conclusion here: All three of these cases involve desire problems that are rooted in belief problems; and, when desire problems are rooted in belief problems, the move to second-order desires is unlikely to help.

*Narrowness vis-à-vis desires for money, desires that flow from addictions, and compulsive desires:* Here the move to second-order desires seems helpful. Most of us desire money, at least from time to time. But most of us know that having money is not

in itself beneficial to us. Thus it makes sense to think that most of us do not desire to desire money. Indeed, what most of us want from ourselves is that we pursue money not for itself, but rather solely as a means to the fulfilling of our desires for things like health and pleasure. Further, it stands to reason that at least the vast majority of smokers desire that they not have the desire to smoke. And the same holds for compulsive desires: We know that those with compulsions desire to be rid of their compulsive desires.

There is, however, an important caveat when it comes to the money case, for there are some people who do not desire to be rid of their desire to have a good deal of money. Donald Trump, for example, desires to be incredibly rich; and, because being rich is so strongly tied to his sense of self-worth, it seems false to say that he desires that he not desire to be rich. Trump differs from most of us. Most of us know that having money is not intrinsically beneficial to us, and so most of us want to be rid of our desire for money. It isn't that we are necessarily ashamed of (or necessarily feel guilty about) the fact that we desire money. Rather, it is just that we know that it is irrational to desire money. We know that rationality requires that we be motivated to pursue money solely as a means. With Trump, however, matters seem very different. He believes that being rich is in itself good for him (and, indeed, for anyone). Thus he doesn't just have desire problems when it comes to money. He has belief problems too. Once belief problems are in play in this way, moving to second-order desires isn't going to help DF theorists make their theories come out right (where 'come out right' means 'come out such that they accord with what we pre-theoretically, or intuitively, think to be true about human well-being').

*Narrowness vis-à-vis bizarre desires:* It may never have occurred to Jarrell that he has the desire to turn radios on whenever he sees them turned off. Indeed, Jarrell simply

may not know that he has this bizarre disposition. In that case, then, Jarrell won't have the desire to be rid of this bizarre desire. However, if Jarrell is aware that he has this bizarre desire, then he may well desire to be rid of this bizarre desire. The general point here, then, is this. When one becomes aware that she has a bizarre desire, she may well form the desire to rid herself of her bizarre desire. But, if one doesn't know about some bizarre desire that she has, then she won't desire to rid herself of this bizarre desire.

*Griffin-style cases (or Dead Sea apple cases):* Dead Sea apple cases usually involve desire problems that are rooted in belief problems. And, when desire problems are rooted in belief problems, the move to second-order desires is unlikely to help. Here consider the case with Dennis. At least before he starts his job at the law firm, he wants the job. His desire for the job is (at least partly) rooted in his defective belief that the job is an attractive job. Is there any reason to think that Dennis desires to be rid of his desire to work at the firm? No, there isn't. The general point here, then, is that, at least with respect to most Dead Sea apple cases, the move to second-order desires will not help.

I am now going to discuss some problems that are peculiar to second-order DF theories. If not for these problems, it would be clear that second-order DF theories are more convincing than first-order DF theories. However, given these problems, it isn't clear that second-order DF theories are more convincing than first-order DF theories.

*Broadness again:* Suppose that Katie desires to have a solid relationship with her nieces and nephews. Also, suppose that she never questions this desire – that she never reflects on this desire. If that is so, then does it really make much sense to think that she desires to have this desire? Why should we think that she has any second-order desire here? Or, put differently, why should we think that she is disposed to bring it about that

she is disposed to bring it about that she has a solid relationship with her nieces and nephews? The truth seems to be that, although Katie has the first-order desire (or disposition) to have a solid relationship with her nieces and nephews, the chain stops right there (so to speak). But, if that is indeed the case, then the second-order DF theory has a broadness problem that first-order DF theories don't have. For surely the state *that Katie has a solid relationship with her nieces and nephews* is an aspect of Katie's well-being. The general worry here, then, is that it may well be that there are a great many aspects of people's welfare that they have first-order desires for, but no second-order desires for.

*The regress problem:* Second-order DF theorists identify someone's welfare with the objects of her second-order desires. But why shouldn't we identify someone's welfare with the objects of her third-order desires, or with the objects of her fourth-order desires, etc.? As Michael Smith says, "each such identification looks to be as plausible as any other. And if each is as plausible as any other then all such identifications are equally implausible. For any identification would require an arbitrary choice between levels."<sup>56</sup> Perhaps this objection can be side-stepped, for perhaps there is no need to pick any particular level. Maybe second-order DF theorists – or, better put, higher-order DF theorists – can claim that we should always go with whatever level of desire happens to be relevant in whatever case is at hand. But then, of course, higher-order DF theorists will have to tell us what 'relevant' means here; and that isn't going to be easy to do.

*Babies:* It seems wrong to think that newborns can have second-order desires. But, if newborns can't/don't have any second-order desires, then, according to second-order DF theories, newborns don't have a well-being – which seems false.

*Second-order desires and one's identity:* Second-order DF theorists often claim that our second-order desires are more closely tied to our identities than our first-order desires. This is likely true for the most part. But consider this example. Suppose that Sharon lives a fairly ordinary suburban life, that she has lived this life for years, that she likes it, that she wants to continue living it, and that she is on the whole quite well-suited for it. Also suppose, however, that she occasionally – every long once in awhile – experiences a sense of disaffection with her suburban life. For instance, as she is driving her kids to soccer practice one day (in her mini-van, of course), she stops and somewhat despairingly thinks: ‘Look what I’ve become – I am so *conventional*. And, perhaps worst of all, I *want* to be this way. Sometimes I really wish I didn’t want to be this way, but, then again, I can’t help that I do.’ The point here is this. Sharon occasionally forms the desire that she not desire to live her suburban life. Yet this occasionally formed second-order desire is not more closely tied to Sharon’s identity than her first-order desire (to live her suburban life) is. And, further, it would be wrong to think that it would be intrinsically good for Sharon not to live the suburban life that she is quite well-suited for.

## **12. How do hypothetical DF theories work?**

What do hypothetical DF theories claim? They claim, roughly, that X is an aspect of A’s welfare if and only if – and directly because – A desires X *in some hypothetical setting* (say, one in which she has just been fully and vividly informed with respect to non-evaluative information).

So as to gain some clarity about how hypothetical DF theories work, it may help if we briefly mention some of Brandt’s thoughts on rational desires. Although Brandt is a

welfare hedonist, Railton and other hypothetical DF theorists have been influenced by Brandt's notion of a rational desire.<sup>57</sup> Brandt says that a rational desire is a desire that would survive or be produced by cognitive psychotherapy, while an irrational desire is one that would be extinguished by cognitive psychotherapy.<sup>58</sup> 'Cognitive psychotherapy' here refers to the process of "confronting desires with relevant information, by repeatedly representing it, in an ideally vivid way, and at an appropriate time...".<sup>59</sup> 'Relevant information' here denotes a subset of 'available information', which itself denotes 'all value-free propositions currently accepted as true by natural science and all value-free propositions that can be justified either by publicly accessible evidence (including that given in self-testimony) or by their logical connection to such evidence'.<sup>60</sup> Also, a piece of available information is here 'relevant' if it in some way affects (or impinges upon) an agent's desire-set, that is, upon her becoming aware of it.<sup>61</sup> In explaining how cognitive psychotherapy is supposed to work, Peter Singer provides the following simple example:

[I]f every time I want to smoke, I imagine, in a suitably vivid way, what it would be like to die prematurely of lung cancer, my desire to smoke may disappear. In that case, Brandt would say, my desire to smoke has not survived maximal criticism by facts and logic. This means it was not a rational desire, and when I acted on it, I was not acting rationally.<sup>62</sup>

Summing up here, then, we might say that the main idea behind cognitive psychotherapy is the following one: An agent is supposed to criticize all of the desires in her desire set by subjecting them to a vivid and thorough confrontation with value-free facts and logic.

Notice that it isn't hard to construct a hypothetical DF theory based on Brandt's notion of cognitive psychotherapy. If we were to do this, then our hypothetical DF theory would say that someone's welfare is wholly determined by what she would desire if she were to have just completed cognitive psychotherapy.



I should stress that different hypothetical DF theories work in at least slightly different ways. Some hypothetical DF theorists (for instance, John Rawls) rely on first-order hypothetical desires, while other hypothetical DF theorists (for instance, Railton and Lewis) rely on second-order hypothetical desires.<sup>63</sup> Moreover, different hypothetical DF theorists incorporate different information conditions into their theories. Railton requires that, when agents are in the hypothetical setting, they are to be given all of the value-free facts that there are.<sup>64</sup> Other hypothetical DF theorists hold that a great many value-free facts, but not necessarily all, are needed.<sup>65</sup> Lewis requires that, when agents are in the hypothetical setting, they are not to be fully informed, but rather are to gain (or, at any rate, are to try to gain) full imaginative acquaintance with various putative prudential values.<sup>66</sup> For example, then, Lewis holds (at least roughly) that, in order for someone to try to determine whether living his life “as a free spirit unbound by law, custom, loyalty, or love” is intrinsically good for himself, he must (1) think really hard – or think vividly and thoroughly – about what it would be like to live his life as a free spirit and then (2) try to see whether he desires to desire to live this sort of life.<sup>67</sup>

Perhaps not too much should be made of the fact that Lewis relies not on full information, but rather on full imaginative acquaintance. After all, full imaginative acquaintance may well entail something close to full information. Lewis states:

I grant one case – a common one – in which one does need empirical knowledge in order to gain imaginative acquaintance with a given putative value. It may be ‘given’ in a way that underspecifies it, with the rest of the specification left to be filled in by reference to the ways of the actual world. For instance when I mentioned the life of a free-spirit as a putative value, what I meant – and what you surely took me to mean – was the life of a free spirit in a world like ours. In such cases, a valuer must complete the specification by drawing on his knowledge of the world, else he will not know what he is supposed to imagine. To that extent – and only to that extent, I think – being well-informed is indeed a qualification for his job.<sup>68</sup>

There are a tremendous number of putative prudential values out there; and, further, on DF theories it seems that *anything and everything* is at least up for being a putative prudential value. Moreover, Lewis's theory presumably requires one to try to gain full imaginative acquaintance with every putative prudential value there is. But, if – with respect to every single putative prudential value there is – one needs to have on hand various value-free facts about the ways of the actual world, then it seems inevitable that, by the time one has gone through the process of trying to gain full imaginative acquaintance with all of the various putative prudential values there are, one will have been close to fully informed.

Speaking generally, hypothetical DF theorists tend to embrace full information conditions. Here hypothetical DF theorists seem to have at least the following two things in mind. (1) They seem to be thinking: 'If – and only if – an agent has full information on hand, then she will be able to subject her desires to the fullest possible confrontation with the value-free facts. Really, reflection is not fully critical unless and until it is carried out with a vivid awareness of all of the value-free facts that there are. Thus, if we want a welfare theory with a very serious critical edge to it, then it makes sense to go with a hypothetical DF theory that employs a full information condition, for such a welfare theory will indeed have a very serious critical edge to it.' (2) Another part of the push toward a full information condition seems to derive from worries about sufficient broadness. In order to ensure that an individual really does hypothetically desire everything that is prudentially fit for herself (or everything that truly is an aspect of her well-being), it is natural to think that she must be informed about everything that is out there.

### 13. Hypothetical DF theories and the stability problem

Just as our actual desires change over time, so too our hypothetical desires change over time. Here consider the following case. Nicole is fifteen, and she very much wants to become a physician. However, she hasn't (as of yet) subjected this desire of hers to any sort of criticism. Suppose, then, that at age fifteen Nicole enters the hypothetical setting and (let's say) undergoes cognitive psychotherapy. Her repeated, vivid, and appropriately timed representations of various pieces of relevant information strengthen her desire to become a physician, while her repeated, vivid, and appropriately timed representations of various other pieces of relevant information weaken her desire to become a physician. After therapy ends her desire to become a physician has survived. Six years pass, and Nicole is now twenty-one. She again enters the hypothetical setting and undergoes cognitive psychotherapy. This time it turns out that, after therapy ends, her desire to become a physician has been extinguished. Let's suppose that it was her representation of what the eight or so years of medical school and residency would entail that made the biggest difference on this score. Representing this matter (and what it entails) at age fifteen weakened her desire to become a physician. But doing the same at age twenty-one *tremendously* weakened her desire to become a physician.

Here consider the following response. 'When Nicole undergoes cognitive psychotherapy at age fifteen, she is given (almost) all of the same information that she is given when she undergoes cognitive psychotherapy at age twenty-one. Having (almost) all of the same information in both cases, will she not thereby have (almost) the same vantage point – or see (almost) all of the same things – in both cases? And will she not in turn, then, come to desire (almost) all of the same things in both cases?'

But this response is off target. At ages fifteen and twenty-one Nicole is, during therapy, given (almost) all of the same information. But to be provided with (almost) all of the same information in both cases is not necessarily to be provided with (almost) the same vantage point. For one's vantage point is a function not only of the information that she is given, but also of her own internalization of the information that she is given. And the latter is itself largely a function of her own personality – that is to say, of her own dispositions with respect to acting, believing, feeling, etc. The main point here, then, is this. Because Nicole's personality can (and no doubt does) significantly change between the ages of fifteen and twenty-one, it follows that her hypothetical desires can (and no doubt do) significantly change between the ages of fifteen and twenty-one.

In light of the fact that our hypothetical desires change over time, it seems that moving to a present moment hypothetical DF theory is the best option that is available to hypothetical DF theorists. However, present moment hypothetical DF theories are problematic. They entail that someone's welfare fluctuates *entirely* in accordance with the fluctuations of her hypothetical desires; and that seems to go too far, for it seems that one's welfare isn't subject to this *extreme* sort of flux. Here someone may protest: 'But *hypothetical* desires don't *radically* fluctuate from moment to moment (month to month, etc.) in the same way that *actual* desires do. Our actual desires may press upon us willy-nilly, but our hypothetical desires are desires that we have reflected upon and criticized in light of the facts. Thus our hypothetical desires won't be willy-nilly occurrences – indeed, they will have more stability to them than our actual desires have. The upshot of this is that maybe present moment *hypothetical* DF theories don't have any trouble capturing the stability that seems to attach to human well-being.' It may be true that

hypothetical desires don't radically fluctuate in the same way that (or as much as) actual desires do. But still, it seems that hypothetical desires do substantially fluctuate from moment to moment, and it goes too far to say that someone's welfare fluctuates *entirely* in accordance with the fluctuations of her hypothetical desires.

Go back to the case of the young academic discussed earlier. Her friends and colleagues harshly criticize the initial draft of her book. She reads what they have to say, and in turn she forms the defective belief that her book is a failure. After this, she loses the desire to finish her book (when in fact finishing her book seems to be an aspect of her welfare). Does moving to a hypothetical DF theory help to prevent this problem from arising? What would happen if our young academic were fully and vividly informed with respect to value-free facts? Would being fully and vividly informed in this way prevent her from forming the defective belief that her book is a failure (and, in turn, prevent her from losing her desire to finish her book)? It might. Vivid, well-informed reflection might force her to think more clearly about all of the details of her situation; and, once she does this, she might see her situation in a less distorted way. In turn, then, she may refrain from jumping to the false conclusion that her book is a failure; and this in turn may prevent her from losing her desire to finish the book.

But, all that being said, there are good reasons for thinking that being fully and vividly informed *wouldn't* prevent our young academic from jumping to the false conclusion that her book is a failure. It would be excessively optimistic to think as follows: When people are in the hypothetical setting and so are exercising their powers of reflection and imagination to the hilt, they will attain a critically distant/quasi-timeless perspective on their own lives and so no longer have moments where they are seeing

things in a distorted way and are forming (defective) desires that flow their seeing things in a distorted way.

We can elaborate on this point by noting that most of us are not particularly adept at transcending our own present emotional/appetitive/cognitive states. Daniel Gilbert (a psychologist) has written about this matter. He cites various studies that show that, when we remember the past and imagine the future, we are *deeply* influenced by our own present states.<sup>69</sup> (“When dating couples try to recall what they thought about their romantic partners two months earlier, they tend to remember that they felt then as they feel now.”<sup>70</sup> “[N]ew mothers abandon promising law careers because they are confident that being home with their children will always be a rewarding job...”<sup>71</sup>) Gilbert says:

Each of us is trapped in a place, a time, and a circumstance, and our attempts to use our minds to transcend those boundaries are, more often than not, ineffective... Imagination cannot easily transcend the boundaries of the present... We assume that what we feel as we imagine the future is what we’ll feel when we get there, but in fact, what we feel as we imagine the future is often a response to what’s happening in the present.<sup>72</sup>

It seems, then, that (at least for most of us) it will be difficult to free ourselves from our own present states when we are in the hypothetical setting. This point matters, and we can see that this is so by thinking about our young academic.

Here consider the following sequence of events. She reads the criticisms from her friends and colleagues, she falls into a very low mood, she forms the defective belief that her book is a failure, and she loses her desire to finish the book. Then, immediately after this, she enters the hypothetical setting, where she is given a bunch of value-free facts. She vividly internalizes these facts. But this doesn’t pull her out of her gloomy mood. Indeed, she retains the belief that her book is a failure, and she doesn’t regain the desire to finish the book.

What can hypothetical DF theorists say in response here? Is there any good reason to think that this sequence of events wouldn't play out? Here hypothetical DF theorists may try to push hard on the notion of *vividness*. They may say that this sequence of events wouldn't play out if our young academic were *vividly* to think about herself, her circumstances, and her future. But how much, really, is it going to help to push hard on the notion of *vividness*? When people are in a gloomy state, it is remarkably hard to pull them out of the overly-strong grip of their present (real or merely perceived) misfortunes – indeed, it is remarkably hard to get them to imagine the future in a way that isn't distorted by their own gloomy mood. Telling them a bunch of value-free facts, and then urging them to let these facts *vividly* sink in, may not help all that much. People can be impervious to facts, and this is especially so in the case of currently gloomy people. As Gilbert says: “when we try to overlook, ignore, or set aside our currently gloomy state and make a forecast about how we will feel tomorrow, we find that it's a lot like trying to imagine the taste of marshmallow while chewing liver.”<sup>73</sup>

Stepping back now, the general point that I am trying to make here is this: Present moment DF theories – be they actual or hypothetical – seem problematic. In order to provide us with a sense of what we are looking for when it comes to stability, it may help if we consider OL theories. What we are looking for, I believe, is a welfare theory that captures the intuition that well-being is both fixed and flowing (or, in other words, both stable and dynamic). And OL theories capture this intuition quite well. With respect to our welfare being *fixed*, OL theories succeed, for OL theories entail that accomplishment, friendship, health, etc., are part of one's welfare throughout her life. And, with respect to our welfare being *flowing*, OL theories also succeed, for OL theories entail that what, for

an individual, falls under accomplishment, friendship, health, etc., will change across time – that is, along with changes in her age, her physiology, her personality, her circumstances, etc. Present moment DF theories, be they actual or hypothetical, clearly have no trouble capturing the intuition that well-being has a flow (or a dynamism) to it. But present moment DF theories fail badly when it comes to capturing the intuition that well-being has a fixity (or a stability) to it. The well-being of our young academic, for instance, doesn't fluctuate entirely in accordance with the fluctuations of her own present desires; and this point holds even if the desires in question are *hypothetical* desires.

There is one other point – and it is a very important one – that ought to be made here. Just as present moment *actual* DF theories have a great deal of trouble making sense of the idea of promoting one's own and others' welfare across time (see above, section 7), so too present moment *hypothetical* DF theories have a great deal of trouble making sense of this same idea. Think of the point this way. On OL theories it is at least tolerably clear what we should shoot for when it comes to prudential decision-making over the long run. An individual should keep her eyes fixed on accomplishment, friendship, health, and so on. Naturally, an individual must leave room for various important contingencies here. For instance, what instantiates accomplishment for one right now may well not instantiate accomplishment for one in the future. But still, the process of promoting one's welfare across time is at least tolerably clear on an OL theory. In contrast, on present moment DF theories – be they actual or hypothetical – it is hard to understand how the process of promoting one's welfare across time is supposed to work. What, exactly, should one keep one's eyes focused on: the objects of one's present (actual or hypothetical) desires, the objects of the (actual or hypothetical) desires that one



believes one will have in the future, the objects of those (actual or hypothetical) desires that one has held for a long time now, or what? And, naturally, the problem doesn't just arise for the promotion of one's own well-being across time. The problem also arises for the promotion of others' well-being across time. It isn't clear how any of this is supposed to work on a present moment DF theory; and moving to a present moment hypothetical DF theory doesn't solve this problem. After all, just as our actual desires can and do change over time, so too our hypothetical desires can and do change over time.

#### **14. Hypothetical DF theories and the *good* part of 'good for'**

With respect to the *good* part of 'good for', we must think about broadness and narrowness. In this section we will start with broadness and then move onto narrowness.

*Broadness:* We know from just above (that is, from section 13) that cases such as that of the young academic pose a broadness problem for hypothetical DF theories. Even after having been fully and vividly informed, our young academic may believe, wrongly, that her book is a failure and, in turn, may no longer desire to finish her book (when in fact it seems that finishing her book is an aspect of her well-being).

Still, when it comes to broadness, it should be admitted that hypothetical DF theories do better than actual DF theories. Here consider the bluegrass case that was discussed earlier (see section 8). The agent has no actual desire to listen to bluegrass because he doesn't know about bluegrass music. However, if he is informed about bluegrass music, then he will know about it and, in turn, may well form the desire for it. (Perhaps, while being fully and vividly informed, the agent not only is told about bluegrass, but also listens to some bluegrass.) And, as for the Mediterranean case (see

section 8), we might say: The agent has no actual desire to take a vacation in the Mediterranean, because he wrongly believes that the Mediterranean is an unattractive vacation spot. However, once the agent is fully informed about the Mediterranean, his beliefs may change and, in turn, he may come to desire to take a vacation there.

One more point about broadness: OL theorists sometimes assert that we can imagine agents who, after having been fully and vividly informed, lack the desire for friendship, or health, or whatever (when in fact these things appear to be intrinsically good for these agents). Here, for instance, consider something Brink says:

If the process of idealization [or, roughly put, the process of fully and vividly informing agents] is purely formal or content-neutral, then it must remain a brute and contingent psychological fact whether suitably idealized subjects would care about things we are prepared, on reflection, to think valuable. But this is inadequate, inasmuch as we regard intellectually and emotionally rich lives as unconditionally good and intellectually and emotionally shallow lives as unconditionally bad. For a person with the normal range of intellectual, emotional, and physical capacities, it is a very bad thing to lead a simple and one-dimensional life with no opportunities for intellectual, emotional, and physical challenge or growth. One's life is made worse, not better, if, after informed and ideal deliberation, that is the sort of life to which one aspires.<sup>74</sup>

Later (in Chapter 3) I will argue that, at least in almost all actual/close to actual cases, agents – be they well-informed or not – *just do* desire all of the basic goods (friendship, accomplishment, etc). If I am right about this, then, although hypothetical DF theories are plagued with some broadness problems, hypothetical DF theorists are nonetheless justified in ignoring broadness related objections (from OL theorists such as Brink) that involve fully and vividly informed agents who happen to lack the desire for health, or accomplishment, or whatever. (In Brink's defense, hypothetical DF theorists have a tendency to open themselves up to this line of objection. For instance, in "Desire as Belief II" Lewis repeatedly stresses that desires are contingent occurrences. And he says:

“Someone might have no desire at all for joy, knowledge, or love.”<sup>75</sup> Given that DF theorists say things like this, it is easy to see why an OL theorist such as Brink might consider himself totally justified in advancing objections that involve fully and vividly informed agents who have no desire for friendship, or accomplishment, or whatever.)

*Narrowness:* Most of us already know that desiring money is irrational. But we desire it anyway, at least from time to time. Most smokers are already well aware that the act of smoking is (in itself) unhealthful for them. But they want to smoke anyway. Most people with compulsive desires already know that the objects of their compulsive desires are worthless. But they have these compulsive desires anyway. Thus providing these people – these people who desire money, or who desire to smoke, or who have compulsive desires – with more *information* isn’t going to cause them to lose their desires for these things. However, the *vividness* requirement may be of at least some help here. Think of the point this way. Many smokers are aware that smoking leads to lung cancer, emphysema, etc., but they push these facts out of their minds whenever they light up. Vivid reflection would force these agents to respond to, or to deal with, these facts. And, once these agents are forced to deal with these facts, they may lose their desire to smoke.

On the other hand, though, it seems likely that at least a great many of these agents – that is, these agents who desire money, or smoking, or repetitive hand-washing – will retain their defective desires even after having been fully and vividly informed. The main problem here is desire recalcitrance. Often agents know that their desires are defective, but, try as they may, they can’t get rid of these desires, for these desires are so firmly entrenched within their motivational structures that they simply can’t be eliminated.

Railton and Lewis have made the move to *second-order* hypothetical DF theories, and this move seems to help when it comes to recalcitrant desires (be they for money, or for smoking, or for repetitive hand-washing, or whatever). For example, if A believes that her desire for smoking is a defective desire, then, even if she can't get rid of her desire to smoke, she almost certainly will desire that she not desire to smoke.

The move to *second-order* hypothetical DF theories may also help when it comes to the case of bizarre desires. Here consider Jarrell (see section 9). Brandt has noted (or at least implied) that strange desires that agents form at young ages may not extinguish during cognitive psychotherapy, just because these desires are so firmly entrenched within these agents' motivational structures.<sup>76</sup> In Jarrell's case the desire to turn radios on whenever he sees them turned off is a desire that he forms at a young age. Remember: He gets into the habit of turning on the radio in his bedroom, in order to drown out his parents' fighting, when he is little. So fully and vividly informing Jarrell may not cause him to lose this bizarre desire of his. Yet it is at least plausible to think that, once Jarrell is forced to reflect on this desire of his, he will form the desire to be rid of this desire.

I don't mean to suggest that second-order hypothetical DF theories don't have any problems at all when it comes to desires for money, desires that flow from addictions, compulsive desires, and bizarre desires. Here, for instance, we might ask: Would fully and vividly informing Donald Trump cause him to desire that he no longer desire money? It is doubtful that it would. When he is being fully and vividly informed, Trump will not be told that money has no intrinsic worth, for that would not be a value-free fact. Thus, even if Trump is fully and vividly informed, he will likely retain his defective evaluative beliefs concerning money; and in turn, then, he will likely still desire (and still desire to

desire) to have tons of money. (There is a general point worth noting here: If someone has a desire problem that is rooted in a belief problem, and if fully and vividly informing him won't rid him of his belief problem, then fully and vividly informing him almost certainly won't rid him of his desire problem either.)

Moreover, we should not forget that there are some problems that are peculiar to second-order actual DF theories (see section 11). It may well be that these problems – or at least that some of them – are also problems for second-order hypothetical DF theories.

Let's now turn to Griffin-style cases (or Dead Sea apple cases). Can hypothetical DF theories handle these cases? At least in most Dead Sea apple cases the relevant desire problems are at least partly rooted in belief problems. For instance, Dennis's defective desire to work at the law firm is rooted in his defective belief that the job at the firm is an attractive job. With respect to Dead Sea apple cases, then, it is important to ask whether fully and vividly informing agents will effectively rid them of their belief problems. For instance, with respect to Dennis, we should ask whether fully and vividly informing him (prior to his taking the job at the firm) would cause him to change his beliefs concerning the attractiveness of the job at the firm. In short, would fully and vividly informing Dennis cause him to see that in fact the job at the firm is not an attractive job?

When Dennis is being fully and vividly informed, he will be told various value-free facts concerning his potential boss, concerning the number of hours that people in his potential position have worked in the past, concerning his potential co-workers, concerning the cases that people at the firm have worked on in the past, etc. Once Dennis is made aware of all of the relevant value-free facts, and once he has vividly internalized all of these facts, it may well happen that he forms the belief that the job at the firm is not

an attractive job; and in turn, then, it may well happen that he loses his desire to work at the firm. This is, of course, the result that hypothetical DF theorists are looking for. Yet there is no *guarantee* that hypothetical DF theorists will get this result. One thing to keep in mind here is that the process of fully and vividly informing someone is intended to be a purely causal process: There are no normative/evaluative constraints, or elements, loaded into the process. The agent is supposed to be given a bunch of value-free facts, the agent is supposed to internalize these facts vividly, and then, once this is over, the agent will have various beliefs and desires – many of which may be quite defective from the standpoint of common sense. It is the desires that the agent winds up with – however defective they may be from the standpoint of common sense – that are supposed to determine the agent’s welfare. Because the process is meant to be a purely causal one, it wouldn’t make sense for hypothetical DF theorists to dictate ahead of time that there are certain beliefs and desires – in particular, ones that are defective from the standpoint of common sense – that the agent simply isn’t allowed to end up with once the process is over. In short, it would be illegitimate if hypothetical DF theorists were to do this.

With respect to Dennis, there seems to be no way for hypothetical DF theorists to guarantee that it won’t happen (1) that at t1 Dennis has on hand all of the relevant facts concerning the job at the firm, (2) that at t1 Dennis believes that the job at the firm is an attractive job and that things will go a certain way if he works there, (3) that at t1 Dennis desires to work at the firm (at least in part because he believes that the job at the firm is an attractive job), and (4) that at t2 Dennis is working at the firm and finding, much to his dismay, that things are not going as he had expected them to go. Or, if we want to spell all of this out more fully, we might say the following. At t1 Dennis has on hand all of the

relevant value-free facts – facts concerning his potential boss, facts concerning the number of hours that people in his potential position have worked in the past, etc. While vividly internalizing all of these facts at t1, Dennis naturally puts more weight on some of these facts, and less weight on others. For instance, Dennis puts a good deal of weight on the fact that, in the past, the firm has dealt with many cases of a certain type, a type that Dennis finds interesting; and Dennis puts less weight on the fact that the people who have worked in his potential job position have averaged sixty hour work weeks for the last five years. At t2, then, Dennis is working at the firm and is finding, much to his dismay, that the firm is not working on as many of the types of cases he likes to work on, and also that he is working seventy hour work weeks. It is not as though Dennis didn't foresee these possibilities – indeed, he was aware that things might turn out this way. Still, he had expected things to go differently, and he is now disappointed that things aren't turning out as he had expected them to go.

Perhaps, in spelling all of this out, we might add that Dennis's expectations were reasonable. He didn't make any cognitive errors when he formed his beliefs about how things would go. Given the facts at the time, nobody could have formed expectations more reasonable than the ones that Dennis formed.

Is there any way for hypothetical DF theorists to ensure that the above sequence of events wouldn't play out? More generally, is there any way for hypothetical DF theorists to guarantee that, once Dennis vividly internalizes all of the relevant value-free facts, he will no longer believe that the job at the firm is an attractive job (and so in turn will no longer desire to work at the firm)? Here hypothetical DF theorists might say: 'What we need to do is to require that, when Dennis is in the hypothetical setting, he is

given facts about how things would go for him if he were to take the job at the firm. If Dennis is provided with facts such as these, and if it is thereby made clear to Dennis that he would be miserable if he were to work at the firm, then this would guarantee the result we are looking for – that is to say, this would guarantee that fully and vividly informing Dennis would cause him to lose his belief that working at the firm is an attractive job (and this in turn would cause him to lose his desire to work at the firm).’

Is it legitimate for hypothetical DF theorists to maintain that, when agents are in the hypothetical setting, they are to be given facts about how things would go for them if such-and-such were to happen? Are there any such facts? While it is clear that there are facts about how things have gone in the past, and while it is clear that there are facts about how things are in the present, it is questionable as to whether there are facts about how things would go for someone if such-and-such were to happen – for instance, about how things would go for Dennis if Dennis were to work at the firm. What we might say, then, is this: ‘When agents are in the hypothetical setting, it is legitimate for them to be provided with facts about the present and facts about how things have gone in the past. Then, with these facts on hand, agents can form beliefs about how things would go for themselves if such-and-such were to happen. However, agents in the hypothetical setting cannot legitimately be given facts about how things would go for themselves if such-and-such were to happen, for it isn’t at all clear that there are any such facts.’

David Sobel would disagree with what has just been said. On his view it is clear that there are facts about how things would go for agents if such-and-such were to happen. In a recent article (“Subjectivism and Idealization”) Sobel says that, in providing an agent in the hypothetical setting with more information than she actually has, the main



goal is “to provide her with a more accurate understanding of what the option she is considering would really be like.”<sup>77</sup> And, elaborating on this matter, Sobel states:

This idealization is an attempt to get already existing facts into better focus...[T]he facts it attempts to get into better focus are the nonnormative facts about what it would be like to have various ways one’s life might go be actual. There is a fact of the matter about this. The idealization is an attempt to get us to see such facts clearly.<sup>78</sup>

Here is how Sobel argues. He starts with a simple case – an ice cream case – where (in Sobel’s estimation) it is clear that there are facts about what it would be like for oneself if one were to eat different flavors of ice cream. Then Sobel generalizes from this case – that is to say, Sobel thinks, in light of reflecting on this ice cream case, that he is justified in thinking that there is a general truth to the effect that there are facts of the matter about what it would be like to have various ways one’s life might go be actual. Two questions arise here. (1) Does the ice cream case work? (Is it clear that there are facts about what it would be like for oneself if one were to eat different flavors of ice cream?) (2) Given that the ice cream case works, is Sobel justified in generalizing from this case?

With respect to the first question, there are good reasons for thinking that the ice cream case works. Here consider what Sobel says:

Think about a careful test of which flavor of ice cream one likes best. One might, and experts no doubt do, devise a raft of idealizing procedures to help in this task. One might hold that a person’s palate should be cleansed between each flavor being assessed. Or one might hold that the person should be comparably hungry when tasting each flavor or that the order of presentation of flavors be changed to make sure that things such as the lingering effect of toothpaste on one’s palate is not affecting the flavor of the first taste and so on. All of this is designed to give the agents more accurate information about what it would be like for them to eat this or that ice cream.<sup>79</sup>

We do tend to think that there are facts of the matter as to what it would be like for us if we were to eat various ice cream flavors. Moreover, we do tend to think that, in order to

get clear on these facts, it is best to do away with whatever might interfere with our ability to know these facts. If Bobby were to eat a bunch of super-hot jalapenos and then, right after that, to try some sort of ice cream that he has never tried, and if he were, after *that*, to say to us that it's not clear to himself whether he likes or doesn't like this sort of ice cream, we would not lose our belief that there is a fact of the matter as to whether he likes or doesn't like (or is indifferent toward) this sort of ice cream. What we might say is this: 'Bobby, let's wait a few days – say, till your taste buds have grown back – and then we'll have you eat this sort of ice cream again. Then, as long as the conditions are right (for instance, as long as you haven't been eating jalapenos again), it will become clear to you – and in turn to us – whether you like this sort of ice cream or not.' The general point here, then, is that Sobel is probably right that there are facts of the matter as to what it would be like for us if we were to eat various flavors of ice cream.

Notice, though, that an ice cream flavor is a fairly stable sort of thing – its nature, or what it is intrinsically like, is not subject to any significant flux or indeterminacy. Further, while an individual's palate is subject to some flux (for instance, the child that doesn't like vegetables may well become the adult who does like vegetables), the fact remains that an individual's palate is apt to stay fairly constant, at least for certain stretches of time. Thus there seems to be space for a fairly stable relationship to develop between an ice cream flavor, on the one hand, and someone's palate, on the other hand. It is within this space that we find settled facts as to what it would be like for someone if she were to eat various ice cream flavors. It isn't at all clear, though, that we should think that this ice cream case gives us evidence for a general claim to the effect that there are facts about what it would be like to have various ways one's life might go be actual.

Think of the matter this way. Since the nature of an ice cream flavor isn't subject to any significant flux, and since (we can assume) Bobby's palate has a fairly set nature, there seems to be no problem with the assumption that there is a fact about what it would be like for Bobby if he were to eat such-and-such an ice cream flavor. But now turn to the case with Dennis. Can we isolate any fact about what it would be like for Dennis if he were to work at the firm? In order to be able to isolate any such fact, we must be able to hold Dennis's personality fairly constant, and we also must be able to hold the nature of the job at the firm fairly constant. But, even if we grant that holding Dennis's personality fairly constant isn't a problem, it isn't at all clear that we can hold the nature of the job at the firm fairly constant. I say this because there are so many variables that are in play here: the number of hours that Dennis works, the various cases that Dennis works on, the moods of Dennis's boss and co-workers (and the moods and temperaments of the various clients for whom Dennis works), the pay-cuts and/or pay-raises and/or bonuses that Dennis may receive, the traffic on the way to work each day, the amount of traveling that Dennis is required to do, etc. All of these things seem to bear on and/or to be intrinsic to the nature of Dennis's job at the firm, and thus all of these things seem relevant when it comes to the question of how things would go for Dennis if he were to work at the firm. Yet all of these things are themselves subject to considerable flux – they certainly aren't stable in the way that the nature of an ice cream flavor is stable.

I am deeply skeptical, then, that there is any settled fact about how things would go for Dennis if he were to work at the firm. To be clear: For *many* of the ways that one's life might go, I don't doubt that there are facts about what it would be like for one if one's life were to go these ways. What I doubt is the claim that, for *all or almost all* of

the ways that one's life might go, there are facts about what it would be like for one if one's life were to go these ways.<sup>80</sup>

The upshot of these remarks can be put like this. Suppose (as seems to be correct) that, prior to Dennis's working at the firm, there is no fact as to what it would be like for Dennis if he were to work at the firm. It follows that, prior to Dennis's working at the firm, Dennis – when he is in the hypothetical setting – cannot be told that, if he were to work at the firm, he would be miserable (for this is not a fact). But, if Dennis is not told this when he is in the hypothetical setting, then it is unlikely that he will lose his defective belief that the job at the firm is an attractive job (and in turn it is unlikely that he will lose his defective desire to work at the firm).<sup>81</sup> It follows, then, that hypothetical DF theories probably cannot handle this case.<sup>82</sup> And, naturally, this is but one of many Dead Sea apple cases that, in my estimation, will pose problems for hypothetical DF theories.

### **15. Hypothetical DF theories and the *for* part of 'good for'**

DF theories that rely on actual desires always give a tolerably clear sense to the *for* part of 'good for'. Given that X is an aspect of A's welfare, the actual DF theorist can say that X is *for* A since and insofar as A has a positive attitude (in particular, a desire) toward X. Hypothetical DF theorists want to retain this attractive feature of DF theories; that is, they want to retain the ability always to give a clear sense to the *for* part of 'good for'. And they want, *in addition*, to give a clearer sense to the *good* part of 'good for' than actual DF theorists do. We have seen (in section 14) that hypothetical DF theories have trouble with the *good* part of 'good for'. For instance, they have trouble with Dead Sea apple cases. Perhaps, though, it should be admitted that, all things considered,

hypothetical DF theories do a better job than actual DF theories when it comes to the *good* part of ‘good for’. But, whether this should be admitted or not, the fact remains that hypothetical DF theories are in bad shape when it comes to the *for* part of ‘good for’. Indeed, on this score it is actual DF theories that have the edge.

Here we should bring up prudential internalism. In its most straightforward form prudential internalism is filled out as follows: If X is an aspect of A’s welfare, then it must be true that A has some favorable mental state (or combination of favorable mental states) vis-à-vis X. DF theories that rely on actual desires straightforwardly capture prudential internalism. The same is not true, however, of hypothetical DF theories.

Think of the matter this way. The basic idea behind hypothetical DF theories is (at least roughly) this. ‘Let’s take an individual and put her in a hypothetical setting where she will be fully and vividly informed. As we are fully and vividly informing her, we will hold her non-belief properties – or, perhaps better put, her personality – constant.’<sup>83</sup> Then, when the process of fully and vividly informing her is over, she will have certain desires. These desires wholly determine her well-being.’ As Connie Rosati has noted, though, it is hard to see how we can hold an individual’s personality constant during the process of fully and vividly informing her – for, as a person is being fully and vividly informed, her personality may well undergo various changes.<sup>84</sup> Indeed, once HA (that is to say, hypothetical A) has been fully and vividly informed, her personality may well markedly differ from A’s. However, in that case, prudential internalism won’t be preserved in any straightforward way. After all, once HA and A no longer have the same personality, this opens up the space for there to be cases where (1) HA desires X and thereby (according to hypothetical DF theories) makes X be intrinsically good for A, and

yet (2) A doesn't desire X and, more generally, has no favorable mental states whatsoever toward X.

But is it true that, once HA has been fully and vividly informed, her personality may well markedly differ from A's? Yes, it is. In order to start to see why this is so, we should ask: What are we going to have to do to an agent to make sure that she becomes *vividly* informed?

At one extreme, we might say that, in order for HA to become *vividly* informed, she is going to have *to live out* all sorts of experiences while she is in the hypothetical setting. For instance, we might say that, in order for HA to become *vividly* informed about what it would be like to be a mom, she must live out the experience of having and raising a child. It seems obvious that, if HA lives out all sorts of experiences in the hypothetical setting, then HA's personality may well undergo some significant changes. We know that this is so, of course, because we know that in real life the living out of experiences can and often does significantly change people's personalities. Naturally, though, hypothetical DF theorists don't have to hold that agents must live out all sorts of experiences when they are in the hypothetical setting.

In order to ensure that HA's personality remains identical to A's (that is, as HA is being fully and vividly informed), it seems that the safest move is for hypothetical DF theorist to restrict the process of fully and vividly informing HA such that HA is simply told a bunch of value-free facts (in a dry way) and then urged to internalize these facts in a suitably *vivid* way. Would this sort of restriction ensure that HA's personality would not significantly change as she is being fully and vividly informed? It is likely that there are plenty of agents whose personalities would not significantly change if they were fully

and vividly informed in this relatively restricted way. However, it is also likely that there are plenty of agents whose personalities *would* significantly change if they were fully and vividly informed in this relatively restricted way. Imagine, for instance, that Shirley is an open and trusting person. But also imagine, that unbeknownst to her and her family, her dad has for years been harboring a secret – say, the secret that he is a pedophile, or a mass murderer, or something like that. Moreover, imagine that, unbeknownst to Shirley, her husband has been having an affair for the last two years with another woman. Now suppose that Shirley heads into the hypothetical setting. At the start of the process of being fully and vividly informed, Shirley is an open and trusting person. She stays that way for a little while, but then she is told about her dad’s secret. Her openness and trusting-ness begins to weaken significantly. Then she is told about her husband’s affair. This shatters her. She immediately becomes a guarded and untrusting person.

Admittedly, the case of Shirley seems rather extreme, and it is doubtful, for most people, that their personalities would drastically change if they were fully and vividly informed. That being said, though, it is very plausible to think that, at least for many people, their personalities would change to an appreciable extent if they were fully and vividly informed. And that is the main point that I want us to see here. How much does it hurt DF theories that they have to allow that, once HA is fully and vividly informed, HA’s personality may well differ from A’s? Rosati has argued that this actually seems to hurt hypothetical DF theories quite a bit.<sup>85</sup>

Here Rosati is thinking the following. If (personality-wise) HA is the same person as A (excepting, that is, for the fact that HA knows more than A), then it might make sense to think that HA’s desires are prudentially authoritative for A. However, if

(personality-wise) HA *isn't* the same person as A, then it is hard to see why HA's desires are prudentially authoritative for A.

It will help if we think about the matter as follows. Suppose that HA and A have different personalities, and suppose that, although HA desires X, A doesn't desire X and, more generally, doesn't have any favorable mental states whatsoever toward X. Now suppose that we tell A that she should desire X because HA desires X. And suppose that A says back to us: 'I don't see why I should desire X. What do HA's desires have to do with what is prudentially good *for me*? I readily admit that HA's desires would have prudentially normative force vis-à-vis me if HA's motivational structure/personality were identical to my motivational structure/personality. But in fact our motivational structures/personalities are different. So, really, I have no reason to care about what HA desires.' In response here we might say the following to A: 'You, A, should care about what HA desires because HA has much more information than you do, and also because HA is very much like you. What is more, HA is psychologically continuous with you, and the only motivational/personality differences between you and her are differences that have been brought about by the process of fully and vividly informing her.' Yet here A might respond: 'Look, HA wants something that I don't want and, more generally, that I don't have any favorable mental states whatsoever toward. And now you're telling me that I should want this particular thing (and, more generally, that HA's desires are prudentially authoritative for me) simply because HA is psychologically continuous with me and knows more than I know. But, frankly, that just misses the point. The point is this. HA and I are *different* people (personality-wise), and so it is hard to see why I should think that *her* desires are prudentially authoritative *for me*.'



What can hypothetical DF theorists say to A here? If there were an independent, legitimately authoritative prudential standard that dictates that A should want this particular thing that HA wants but that A doesn't want, then it would obviously follow that A should want this particular thing that HA wants but that A doesn't want. But, naturally, hypothetical DF theorists cannot appeal to any independent standard such as this one. After all, hypothetical DF theories (by their very nature) entail that HA's desires set the *only* standard that has any legitimate prudential authority vis-à-vis A.

Or, if the foregoing isn't totally clear, then it may help to think of the matter in the following way. Here hypothetical DF theorists are caught in a bind because A is questioning the prudential authority of HA vis-à-vis her (that is, vis-à-vis A). So as to get themselves out of this bind, hypothetical DF theorists might want to appeal to some third party to settle the matter – that is to say, hypothetical DF theorists might want to appeal to someone or something that is independent of both A and HA. But this move isn't open to hypothetical DF theorists, for hypothetical DF theories (by their very nature) entail that there is no third party that has any legitimate prudential authority in this context.

Have hypothetical DF theorists addressed this issue? Hypothetical DF theorists have, in general, assumed that HA's personality can be held constant during the process of fully and vividly informing HA. So in laying out their theories hypothetical DF theorists haven't, in general, seen the need to address this issue. It is worth noting, though, that Railton does address this issue – or at least something close to it. Unlike other hypothetical DF theorists, Railton is explicit that, when HA is forming her desires, she must firmly keep in mind that she is about to assume the place of A.<sup>86</sup> In this regard Railton says that the wants of HA that determine A's well-being “are wants regarding

what he [HA] would seek were he to assume the place of his actual, incompletely informed and imperfectly rational self, taking into account the changes that self is capable of, the costs of those changes, and so on.”<sup>87</sup> And Railton adds:

A fully informed and rational individual would, for example, have no use or desire for psychological strategies suited to circumstances of limited knowledge and rationality; but he no doubt would want his incompletely informed and imperfectly rational actual self to develop and deploy such strategies.<sup>88</sup>

I don’t think that this sort of move solves the problem, though.<sup>89</sup> As Rosati notes:

The fully informed person is, of course, supposed to determine what she would want the actual person to want, were she about to assume the place of that person. But this fact does not insure that the fully informed person will react as someone the actual person might regard as an ideal advisor, since it does not insure that she has the traits the actual person would seek or ought to seek in an ideal advisor.<sup>90</sup>

So far as I can tell, Rosati is right. Once HA and A come to have different motivational structures/personalities, it is hard to see why A should care about what HA desires – or, in other words, it is hard to see why HA’s desires are prudentially authoritative for A. And this is so even with respect to Railton’s theory, which says that, when HA is forming her desires, she must firmly keep in mind that she is about to assume the place of A.

To conclude Chapter 1: Actual DF theories struggle mightily with stability and with the *good* part of ‘good for’. In relation to stability and the *good* part of ‘good for’, it may be true (for all I know) that hypothetical DF theories do better than actual DF theories. Still, there is no denying that hypothetical DF theories seriously struggle in relation to stability and the *good* part of ‘good for’. Also, hypothetical DF theories have a key problem that actual DF theories don’t have: Hypothetical DF theories have trouble with the *for* part of ‘good for’.<sup>91</sup> In Chapter 2 we will look at OL theories. They do well with stability and with the *good* part of ‘good for’. However, OL theories struggle with the *for* part of ‘good for’. Since this is so, we should try to do better than OL theories.

## Chapter Two: Objective List Theories

### 16. Two common objectivist moves

The main argument of this chapter can be put as follows. In order adequately to capture the *for* part of ‘good for’, a theory of human welfare must embrace a relatively strong form of prudential internalism. OL theories fail on this score (though see section 18 for a caveat). Therefore we should be strongly inclined to reject OL theories. The main argument of this chapter will come in sections 19-21. Here in section 16, and also in sections 17 and 18, we will mostly just be setting the stage.

There are two moves that objectivists about human well-being commonly make: the move to a list account and the move to a perfectionist account. Here in section 16 we will discuss these two moves. We can start with the move to a list account.

Suppose that someone says: ‘I accept objectivism about human welfare. After all, there are plenty of things that benefit one even if one has no favorable attitude toward them. For example, friendship, health, and accomplishment seem to be part of one’s welfare irrespective of whether one has any favorable attitude toward them. Still, I have no *theory* of human welfare. For I don’t have any particularly informative answer to give to the question of what directly makes something (anything) be an aspect of someone’s welfare.’ The move to a list account seems to help here. Once an objectivist advances a list of basic goods (or, perhaps better put, a specific interpretation of what is on *the* list of basic goods), she can appeal to the list as a way of helping her to explain what directly makes something (anything) be an aspect of A’s welfare. Suppose that A is writing a great novel, and also that some OL theorist holds that accomplishment is one of the basic goods. If we ask this OL theorist *why* A’s writing a great novel is an aspect of A’s

welfare, this OL theorist can answer: ‘A’s writing a great novel is an aspect of A’s welfare *because* A’s writing a great novel instantiates one of the basic goods for A, namely, the basic good of accomplishment.’

Of course, even if an objectivist moves to a list account, we still might wonder whether she has provided us with a *theory* of human well-being. Sumner claims that merely providing us with a list of basic goods does not amount to a *theory* of human welfare.<sup>92</sup> He says: “[A] list of virtues is not a theory of virtue, and a list of (human) goods is not a theory of welfare.”<sup>93</sup> Moreover, with respect to this same issue, Brink says that, if an OL theorist advances “a mere list of goods, with no unifying strands, it begins to look like a disorganized *heap* of goods.”<sup>94</sup> And Shelly Kagan also discusses this matter. He says:

What is it, if anything, that unifies the various items on the list? Why do these things, and only these things, have objective value? Not surprisingly, answers to these questions are difficult to come by, and different objectivists offer different answers. Many proposals take the form of *perfectionism*: the objective goods are those that are elements in an ideal or perfect human life, one which fully realizes the distinctive and essential characteristics of human nature. Of course, even within a broadly perfectionist approach, there is room for considerable disagreement, both with regard to what constitutes human nature and – accordingly – what goods are elements in the perfect human life.<sup>95</sup>

We might say that there are two very general sorts of OL theories: *neutral* OL theories and *perfectionist* OL theories. Neutral OL theories claim that the basic goods are prudential goods, but they go no farther – that is, neutral OL theories are neutral about whether the basic goods are perfectionist goods. Perfectionist OL theories claim that the basic goods are both prudential goods and perfectionist goods.

How do neutral OL theories work? Neutral OL theorists advance a list of basic goods (or, perhaps better put, a specific interpretation of what is on *the* list of basic

goods). And neutral OL theorists claim that X is an aspect of A's welfare if and only if (and directly because) X instantiates one of the basic goods for A. For example, if X instantiates accomplishment for A, it follows on neutral OL theories that X is an aspect of A's welfare (and also that X is so directly because X instantiates accomplishment for A). If we push the neutral OL theorist to say more – if we ask her 'But what unifies the basic goods such that they form one unit, as opposed to being a mere heap of disconnected goods?' – then the neutral OL theorist will answer: 'There may be a unifier here, but, whether there is or not, I will leave that open.'

Like neutral OL theorists, perfectionist OL theorists advance lists of basic goods; and, like neutral OL theorists, perfectionist OL theorists claim that X is an aspect of A's welfare if and only if (and directly because) X instantiates one of the basic goods for A. However, if you ask a perfectionist OL theorist what unifies the basic goods, then the perfectionist OL theorist will give you an answer: The perfectionist OL theorist will claim that the basic goods are unified in that they are all perfectionist goods. The idea behind something's being a perfectionist good for someone is that it directly contributes to the proper/excellent functioning of her *qua* her being the sort of thing she is, namely, a human being. Thus we might say: X is pro tanto perfective of A if and only if X directly contributes to the proper/excellent functioning of A as a human being. (Naturally, perfectionist goods – like prudential goods – need not be goods merely for humans. There are goods that directly contribute to the proper/excellent functioning of golden retrievers, and such goods are perfectionist goods for golden retrievers. Likewise, there are goods that directly contribute to the proper/excellent functioning of tulips, and such goods are perfectionist goods for tulips. And, of course, the set of perfectionist goods for

humans differs from, though perhaps overlaps with, the set of perfectionist goods for golden retrievers, which itself differs from, though perhaps overlaps with, the set of perfectionist goods for tulips.)

In spelling out their view there are two routes that perfectionist OL theorists might take. (1) *The identity route*: Perfectionist OL theorists might say that prudential value and perfectionist value are identical. (2) *The non-identity route*: Perfectionist OL theorists might deny that prudential value and perfectionist value are identical.

The first of the above two routes – the identity route – can be filled out roughly as follows. There are various desire-independent, relational goods: friendship, health, accomplishment, knowledge, and so on. These goods are both prudential goods and perfectionist goods. Whenever someone gets an instance of one of these goods for herself, she will thereby be made better off in prudential terms; and, moreover, whenever someone gets an instance of one of these goods for herself, her own human nature (and, by extension, herself taken as a whole) will thereby be pro tanto perfected.<sup>96</sup> Moreover, the really important claim here is this: The furthering of A's well-being and the pro tanto perfecting of A's own human nature are *one and the same thing* (or are identical to each other). Put another way, then, the really important claim here is this: A's faring well just is A's functioning properly/excellently (or well) as the sort of thing she is, namely, a human being. (Perfectionist OL theorists who take the identity route might allow that there is a conceptual split between prudential value and perfectionist value. On their view, though, prudential value and perfectionist value are identical at least in the sense that the same thing answers to these two concepts. Think of water and H<sub>2</sub>O here: These may be different concepts, but the same thing answers to these two concepts.)

Murphy is a perfectionist OL theorist who takes the identity route. As Murphy says: “on my view, prudential value just is perfectionist value.”<sup>97</sup> And there are other perfectionist OL theorists who also seem to take the identity route – for instance, John Finnis and Richard Kraut.<sup>98</sup>

As for the non-identity route, I can’t, with certainty, name anyone who takes this route. I strongly suspect, though, that there are some OL theorists who at least implicitly accept this view. Here is one way to spell out the non-identity view: Prudential value and perfectionist value are not identical, but it turns out (1) that everything that is prudentially good for one is also perfective of one and (2) that everything that is perfective of one is also prudentially good for one. And here is another way to spell out the non-identity view: It might be said that, although everything that is prudentially good for one is also perfective of one, not everything that is perfective of one is prudentially good for one.

To sum up: All OL theorists conceive of the basic goods as prudential goods. Neutral OL theorists take a neutral stance on the question of whether the basic goods are not only prudential goods, but also perfectionist goods. All perfectionist OL theorists claim that the basic goods are not only prudential goods, but also perfectionist goods. Some perfectionist OL theorists hold that prudential value and perfectionist value are identical to each other. Other perfectionist OL theorists deny that prudential value and perfectionist value are identical to each other.

## **17. The main strengths of OL theories**

In discussing the main strengths of OL theories, we should focus on stability and the *good* part of ‘good for’. Here, then, let’s start with stability.

I briefly discussed OL theories with reference to stability at the end of section 13. Here I will briefly elaborate on what I said there. Consider accomplishment. According to OL theorists, accomplishment is part of one's welfare throughout one's life. There is no flux (or instability) here: It is not as though accomplishment is part of one's welfare at one moment, but then not a part of one's welfare at some later moment. And this seems to get the matter right. After all, we do not tend to think of accomplishment as something that is sometimes prudentially good for us, and sometimes not prudentially good for us. On the contrary, we tend to think of accomplishment as always being prudentially good for us. Or, to be more precise, we tend to think of accomplishment as always being a near-possible component of our well-being. That is, we tend to think that, for any given human, it is always true that engaging in accomplishment is beneficial for her, and that this is so irrespective of whether or not she is currently engaging in accomplishment.

Another thing to note here is that OL theories operate on two levels: the level of the basic goods and the level of the instances of the basic goods. While accomplishment is always prudentially good for A (or is prudentially good for A throughout her life), what instantiates accomplishment for A will change across time. This is a nice feature of OL theories: We must allow that the constituents of someone's welfare change across time – along with changes in her age, physiology, circumstances, abilities, etc. When A is very little, learning how to talk and how to walk will instantiate accomplishment for her. As she grows older, these things will no longer instantiate accomplishment for her: New things will instantiate accomplishment for her (say, learning her ABC's and coloring within the lines). And, as she grows older still, more changes in this regard will follow. All the while, though, accomplishment will remain as a fixed part of her welfare.



Speaking generally, then, OL theories seem to handle stability quite well. I say this because we do tend to think of our well-being as both fixed and flowing (or as both stable and dynamic); and OL theories have no trouble capturing this intuition of ours.

Now let's turn to OL theories in relation to the *good* part of 'good for'. Here there are two things to keep in mind: narrowness and broadness.

*Narrowness*: Here it may help if we briefly consider some of the cases that we brought up in Chapter 1. Suppose that Jarrell sometimes flips radios on when he sees them turned off. On OL theories this activity does not enter Jarrell's welfare, for this activity doesn't instantiate any of the basic goods for Jarrell. (One might say: 'But perhaps this is an accomplishment of some sort – after all, Jarrell is definitely doing something when he flips radios on.' Naturally, though, if we are using the OL theory sense of 'accomplishment', which is normatively/evaluatively loaded, then it is clear that this is not an accomplishment.<sup>99</sup>) Another case: Suppose that things go badly for Dennis when he works at the law firm. Here OL theorists will readily admit that certain *aspects* of Dennis's experience at the firm add to his welfare. For instance, if Dennis works on a few interesting cases during his time at the firm, then OL theorists will say that Dennis's doing so at least to some extent instantiates accomplishment, or knowledge, or pleasure (or some combination of these) for himself. What OL theorists will not admit, though, is that the state *Dennis's working at the law firm* (taken as a unit that encompasses Dennis's entire experience of working at the firm) adds to Dennis's welfare. After all, this state (taken as a whole) doesn't in fact instantiate any of the basic goods for Dennis – again, only certain portions of Dennis's experience of working at the firm fall under the basic goods for Dennis, and so on OL theories only certain portions of Dennis's experience of

working at the firm count as aspects of his welfare. (In order to give a full account of what is going on here, OL theorists would need to provide us with a worked-out account of ill-being. But, that said, the main point here still stands: OL theorists need not say what seems false, namely, that Dennis's working at the firm adds to his welfare.)

In relation to OL theories and narrowness, one might say: 'Friends sometimes hurt each other, and in fact certain friendships are harmful to people. So it is implausible to think that friendship always benefits people.' But this is a misunderstanding. When friends hurt each other, the basic good of friendship is not being instantiated for anyone. What is on display in these cases is not friendship – rather, it is meanness, or insecurity, or resentment, or whatever. And, *mutatis mutandis*, the same holds for the other basic goods.<sup>100</sup>

*Broadness:* In relation to OL theories and broadness, one might say the following. 'The list structure assumes that all humans are benefited by the same goods. However, someone might benefit from some rather uncommon activity – say, mountain-climbing, or bird-watching. And, because of the uncommonness of the activity in question, the activity in question may not be the sort of thing that can be captured by the list structure – in which case it follows that OL theories are insufficiently broad, since they are failing to include within the content of some agent's welfare everything that truly is an aspect of his welfare.'<sup>101</sup>

But this objection fails. Activities such as mountain-climbing, playing the piano, bird-watching, etc., can and obviously do instantiate one or more of the basic goods for certain people (depending on their circumstances, abilities, life-histories, and so on). After all, these sorts of activities can and obviously do instantiate accomplishment, or

health, or aesthetic experience, or knowledge (or some combination of these) for certain people. Also, if activities such as these bring someone pleasure, then (according to those OL theorists who hold that pleasure is a basic good) the pleasure that these activities bring to this person adds to her welfare.

Griffin captures the general point that I am making here when he says: “There being just one profile of prudential values for humans [or one list of basic goods for humans] is compatible with there being very many forms that a good human life could take...Monism about profiles is compatible...with all acceptable forms of pluralism.”<sup>102</sup> This seems right. For OL theories, there is no worry that the list structure can’t capture all of the things that truly are prudentially good for agents – or, in other words, there is no worry about insufficient broadness.

Unlike DF theories, OL theories have no trouble with the *good* part of ‘good for’. This obviously counts in favor of OL theories.

## **18. Non-standard OL theories and prudential internalism**

There aren’t any contemporary OL theorists who accept a relatively strong form of prudential internalism. But is it possible for an OL theorist consistently to accept a relatively strong form of prudential internalism? This does seem possible.

We can distinguish between *standard* OL theories (which reject all relatively strong forms of prudential internalism) and *non-standard* OL theories (which accept a relatively strong form of prudential internalism). Moreover, we can distinguish between two types of straightforward prudential internalism. *Constitutive* straightforward prudential internalism says that, if X is an aspect of A’s welfare, then A must have some

sort of pro-attitude toward X *and* this pro-attitude has to be at least part of whatever it is that directly makes X be an aspect of A's welfare. This is the version of prudential internalism that actual DF theorists accept. *Non-constitutive* straightforward prudential internalism says that, if X is an aspect of A's welfare, then A must have some sort of pro-attitude toward X *but* this pro-attitude cannot be any part of whatever it is that directly makes X be an aspect of A's welfare. Non-standard OL theories always accept something that is at least very close to non-constitutive straightforward prudential internalism. I am aware of two sorts of non-standard OL theories: the epistemic contact sort and the enjoyment sort.

*Epistemic contact non-standard OL theories:* Darwall has discussed views that are similar to epistemic contact non-standard OL theories. He states:

Existence internalism is a metaphysical claim. It is the nature of unqualified normativity, it holds, to be necessarily related to motivation. But this general characterization admits of two very different kinds of view. On one, the ethical is such that motivation is necessarily an effect of engaging it epistemically. On the other, motivation is a constituent of ethical facts themselves. The most familiar versions of the former are rational intuitionisms such as Plato's which hold that it is the nature of the Good that it cannot be fully grasped with indifference. Richard Price held a similar view about moral obligation: "When we are conscious that an action...*ought* to be done, it is not conceivable that we can remain *uninfluenced*, or want a *motive* to action." These may seem to be versions of judgment internalism, but they are not. They say, not that ethical belief or sincere assertion necessarily motivates, but that actual consciousness of or cognitive contact with the ethical does. But Price also stresses that motivation has nothing to do with what the ethical facts themselves are. Motive is "the *effect* of obligation perceived, rather than *obligation itself*."<sup>103</sup>

The point that I want to make here is this. Epistemic contact non-standard OL theories claim that, if A comes into epistemic contact with something (anything) that is an aspect of her welfare, then she inevitably will (at that very moment) form a desire for this thing – though, of course, her desire won't at all make this thing be an aspect of her welfare.

*Enjoyment non-standard OL theories:* An enjoyment non-standard OL theorist will lay her view out as follows. ‘If and when A gets some (any) actual aspect of her own welfare, then she inevitably will (at that very moment) have the pro-attitude of enjoyment toward this state of affairs. Her enjoyment of this state of affairs will not at all make this actual aspect of her welfare be an aspect of her welfare. But still, her enjoyment of this actual aspect of her welfare will follow upon her getting this actual aspect of her welfare, and inevitably so.’ Fred Feldman has mentioned (at least roughly) this sort of non-standard OL theory. He rejects OL theories, but he mentions that he “would be prepared to look more closely into” OL theories that “could somehow guarantee that the items on the list would be things that would simply *have to be enjoyed by the one who had them.*”<sup>104</sup>

Also, Christine Korsgaard may incline toward enjoyment non-standard OL theories. In the final endnote of her article “The Myth of Egoism”, she states:

[I]f we reject both hedonism and the desire-satisfaction model, it may seem as if we are left with only a kind of “external realist” conception of the individual’s good or happiness. On such a conception, the good is something defined independently of the individual’s natural desires and capacities for interest and enjoyment. This seems absurd, since most of us believe that a person’s good or happiness must be something *necessarily* capable of motivating, interesting, or pleasing him...The ancient Greeks, especially Aristotle, offer a conception of the human good which is psychologically grounded, but which cannot be identified with either the desire-satisfaction model or Benthamite hedonism. The rough idea is that happiness rests in the excellent activity of our healthy faculties, an activity that we necessarily experience as pleasurable, although not because it is the cause of a pleasant sensation. I believe some version of this conception can be shown to be much more plausible than its modern, less sophisticated, alternatives.<sup>105</sup>

Here, then, Korsgaard may be pushing an enjoyment non-standard OL theory – or, in other words, she may be pushing a form of objectivism that says that, although A can’t help but enjoy every actual aspect of her welfare, none of the aspects of A’s welfare are

even in part directly constituted by A's enjoying them. (I have said '*may* be pushing' because, honestly, I am not sure exactly what Korsgaard's view of well-being is.)

If we set Korsgaard's yet-to-be-worked-out view to the side, it seems safe to say that there aren't any contemporary OL theorists who accept non-standard OL theories. In sections 19-21 my argument will be that OL theories are implausible since they don't accept any relatively strong form of prudential internalism. But my target here is really just all currently circulating OL theories: I won't argue against non-standard OL theories.

The major difference between non-standard OL theories and my own theory – the no priority theory – is this. Like all OL theories, non-standard OL theories say that X is an aspect of A's welfare if and only if (and directly because) X instantiates one of the basic goods for A. Thus, if some state of affairs is an aspect of A's welfare, then, in explaining why this state of affairs is an aspect of A's welfare, non-standard OL theorists need not reference any pro-attitude of A's – they need only say that this state of affairs is an aspect of A's welfare directly because this state of affairs instantiates one of the basic goods for A. The main point here, then, is that non-standard OL theorists join standard OL theorists in denying that A's pro-attitudes do any of the direct constitutive work when it comes to A's welfare – that is to say, non-standard OL theorists join standard OL theorists in denying that A's pro-attitudes directly make states of affairs be aspects of A's welfare.<sup>106</sup> The no priority theory is different. On the no priority theory A's desires do direct constitutive work vis-à-vis A's welfare. This is so because on the no priority theory A's desires are built into the truth conditions for something's (anything's) being an aspect of A's welfare. As I noted in section 2, the no priority theory's central claim is this: X is an aspect of A's welfare if and only if – and directly because – (1) X

instantiates friendship, or accomplishment, or health, or whatever for A and (2) X's obtaining satisfies some desire that A has.

There is a reason, I think, for why there are no contemporary OL theorists who officially (or explicitly) have worked out a non-standard OL theory. The reason is that non-standard OL theories are not particularly attractive. Here consider enjoyment non-standard OL theories. They say: 'If X is an aspect of your welfare, then X is so *whether you enjoy it or not* and, more generally, *whether you have any sort of pro-attitude toward it or not* – by the way, though, if you do happen to get X, then you necessarily will enjoy it.' I don't see any *clear* – or any *obvious* – incoherence here. But this is a strange claim. After all, it is strange to say: 'This thing is prudentially good for you *whether you enjoy it or not* – but still, there definitely is a necessary connection between this thing's being prudentially good for you and your enjoying it.' (The problem for epistemic contact non-standard OL theories is the same. They seem strange in that they say: 'If X is an aspect of A's welfare, then X is so *whether A desires X or not* and, more generally, *whether A has any sort of pro-attitude toward X or not* – by the way, though, if A does happen to come into epistemic contact with X, then A necessarily will, at that moment, desire X.')

I think that, if we are going to hold that prudential value for humans necessarily links to the pro-attitudes of humans, then we shouldn't be obscure or tentative about what we are doing. In other words, I think that we should just go ahead and claim that the necessary link in question is a *constitutive* one.

Having said all of that, I am not certain that non-standard OL theories are totally unworkable. And, moreover, I must admit that, all things considered, non-standard OL theories are quite similar to my own theory (the no priority theory). After all, both non-

standard OL theories and the no priority theory rely on the same desire-independent, relational goods: friendship, aesthetic experience, knowledge, health, and so on (though it should here be said that these two theories conceive of these goods differently – non-standard OL theories conceive of these goods as *prudential* goods, while the no priority theory doesn't). Also, both non-standard OL theories and the no priority theory ensure that there is at least a fairly tight link between the aspects of one's welfare and one's own pro-attitudes. Maybe, then, the differences between these two theories are not that important in the grand scheme of things. Still, if you ask me which is better, I will tell you that the no priority theory is better than non-standard OL theories. Why, after all, should we claim (1) that X is an aspect of A's welfare irrespective of whether A has any sort of pro-attitude toward X and then proceed also to claim (2) that there is some sort of necessary connection between X and A's pro-attitudes? This isn't an attractive position.

In any case, I will not be arguing against non-standard OL theories in the next few sections – I will only be arguing against standard OL theories. Henceforth, let's refer to standard OL theories as OL theories; and let's now go ahead and see why these theories have trouble with prudential internalism and, by extension, with the *for* part of 'good for'.

### **19. Prudential internalism: what it says and doesn't say, and why it seems true**

In its most general form prudential internalism says that, if X is an aspect of A's welfare, then there must be some sort of a positive linkage/connection/hookup between A and X. Further, in its most straightforward form prudential internalism says that, if X is an aspect of A's welfare, then A must have some sort of pro-attitude toward X. More than anything else, it is *alienation* that is central here: We are looking for an account of



human well-being that is sufficiently non-alienating; that is to say, we are looking for an account of human well-being that ensures that every aspect of A's welfare is truly *for A*. Remember: It is one thing to say that X is intrinsically good – it is another thing to say that X is intrinsically good *for A*. Something can be intrinsically good, and yet have nothing to do with A. Our concern here is with what is intrinsically good *for A*.

Prudential internalism is a certain type of (what Darwall refers to as) existence internalism. Darwall notes, rightly, that we should make sure to distinguish existence internalism from judgment internalism. He puts the matter as follows:

[T]he range of positions labeled internalist in ethical writing is bewilderingly large, and only infrequently are important distinctions kept clear. Sometimes writers have in mind the view that sincere *assent* to a moral (or, more generally, an ethical) judgment concerning what one should do is necessarily connected to motivation (actual or dispositional). This necessity may be conceptual, or perhaps metaphysical, the thought being that it is not merely a contingent matter that people have motives to do what they think or sincerely say they should. I call internalism of this variety *judgment internalism* and distinguish it from another set of theses that concern, not what it is to accept an ethical judgment, but what it is for such a judgment to be true. According to *existence internalism*, someone morally (or ethically) ought to do something only if, necessarily, she (*the agent*) has (actually or dispositionally) motives to do so. Again, this necessity might be conceptual or metaphysical, the thought being that it is not merely a contingent matter that agents have motives to act as they ought.<sup>107</sup>

Judgment internalism figures centrally in the debate between cognitivists and non-cognitivists. For our purposes here, though, judgment internalism isn't directly relevant – only existence internalism is. I say this because prudential internalism is a type of existence internalism. (Note that existence internalism might concern values or reasons. Prudential internalism is a thesis about prudential values. It is not directly a thesis about prudential reasons. However, it makes sense to think that, if X is an aspect of A's welfare, then A has a pro tanto prudentially normative reason to make X obtain. Thus it

makes sense to think that, although prudential internalism is not directly about prudential reasons, it nonetheless is indirectly about prudential reasons.)

Existence internalism is often advanced as a thesis about normativity in general/unqualified normativity. For instance, in his well-known article “Internal and external reasons” Bernard Williams argues for a version of existence internalism that concerns normativity in general/unqualified normativity.<sup>108</sup> Williams argues that A can’t have a pro tanto unqualifiedly normative reason to make X obtain unless A is (actually or dispositionally) motivated to make X obtain. Those who accept prudential internalism need not accept Williams’s position. And that is a good thing, for it seems implausible to claim (roughly) that every normative reason that someone has must positively hook into her own subjective motivational set. On the other hand, prudential internalism seems quite level-headed. Indeed, it seems eminently plausible to claim (roughly) that every *prudentially* normative reason that someone has must positively hook into her own subjective motivational set. Think of the matter this way. Existence internalism has an air of strong *partiality* about it: This thesis connects normative reasons to an individual’s motives such that something is a normative reason for someone only if she herself is (actually or dispositionally) motivated by this thing. It is sensible to think that existence internalism holds true when we are talking about nothing but the realm of prudence. After all, the realm of prudence is the realm of self-interest. It is, then, a realm of strong *partiality*. But, once we are talking about a different realm – say, the realm of morality (which seems to be a realm of strong *impartiality* and/or strong *constraint*), or the realm of normativity in general/unqualified normativity (which doesn’t seem to reduce to nothing but the realm of prudence) – existence internalism looks far less plausible.

Some comments concerning the relationship between perfectionist value and prudential internalism are in order here. I often say things like this: It is one thing to say that X is intrinsically good – it is another thing to say that X is intrinsically good *for A*. When I say things like this, I am trying to push us toward a relatively strong form of prudential internalism – that is, I am trying to get us to admit that the *for* part of ‘good for’ can’t properly be captured unless we accept a relatively strong form of prudential internalism. However, I am willing to accept that perfectionist goods are intrinsically good for agents, and that they are so in a desire-independent way (and, more generally, in a pro-attitude independent way). There is, then, a question that OL theorists might have for me: Why do I think that we must accept prudential internalism, when I already admit that there is a sense of ‘intrinsically good for A’ that makes no reference at all to A’s pro-attitudes? What is the difference between the *for* part of ‘good for’ in case of prudence (on the one hand) and the *for* part of ‘good for’ in the case of perfectionism (on the other hand)? Why is it that in the former case I hold that a reference to A’s desires is needed, while in the latter case I hold that no reference to A’s pro-attitudes is needed?

My answer, in a nutshell, is this: Prudential goods are *personal* in a way that perfectionist goods are not. Perfectionist values certainly are relational values. If X is intrinsically good for me in the perfectionist sense, then X is a value for me. But here we need to ask: In what *way* are perfectionist goods for their recipients? More specifically, we should ask: Are perfectionist goods for their recipients *qua* their being the *individuals* they are? Or, instead, are perfectionist goods for their recipients *qua* their being the *kind* of thing they are? The answer is that perfectionist goods are for their recipients *qua* their being the *kind* of thing they are. If X is intrinsically good for me in the perfectionist

sense, then X is intrinsically good for me *qua* my being the *kind* of thing I am, namely, a human being. Matters are different with prudential values, for prudential values are for their recipients *qua* their being the *individuals* that they are.

Because prudential values are *personal* in a way that perfectionist values are not, it is sensible to think that the *for* part of ‘good for’ in the case of prudential value must be filled out differently than the *for* part of ‘good for’ in the case of perfectionist value. In particular, it is sensible to think that a reference to A’s own mental states/pro-attitudes is necessary when it comes to the spelling out of ‘good *for* A in the prudential sense’, though not when it comes to the spelling out of ‘good *for* A in the perfectionist sense’.

Various contemporary philosophers have suggested that there is something especially *personal* about prudential value that distinguishes it from perfectionist value. For instance, Sumner has tried to bring out this point out, and, although I disagree with much of what Sumner says about this matter, it will help to quote him. Sumner says:

What distinguishes welfare from all other modes of value is its reference to the proprietor of the life in question: although your life may be going well in many respects, it is prudentially valuable only if it is going well *for you*. This subject-relativity is an essential feature of our ordinary concept of welfare. It does not merely rest on the truism that all welfare is someone’s welfare, the welfare of some particular individual. If lives are the sorts of things that can have perfectionist value then personal excellence is always the excellence of some particular person, but the category of perfectionist value is free of the relativizing indexicals which are characteristic of well-being. Among the modes of value which can belong to individual lives, welfare stands out by virtue of incorporating an internal reference to its bearer.<sup>109</sup>

What Sumner should say here, I believe, is something like this. Prudential values are for the agent, taken as the individual she is (full stop). In contrast, perfectionist values are for the agent, taken as the *sort* of thing she is, namely, a human being. (Here is another way to put the difference. Prudential values target the person. In contrast, perfectionist

values target the person's own human nature and then, after that, end up flowing to the person as a whole. Or, put differently, perfectionist values pro tanto perfect the person's own human nature and then, by extension, pro tanto perfect the person as a whole.)

Sumner tends to think of perfectionist value as a sort of value that, though it is *for one*, is nonetheless for one in a way that is somehow or other *external*. With respect to a life's going well in perfectionist terms, Sumner says: "it just does not seem true that my life automatically goes better *for me* if the goals I am pursuing rank higher rather than lower from this external standpoint."<sup>110</sup> To think that perfectionist value is an *external* value is, to some extent, a distortion. After all, it's not as though perfectionist value perfects anything outside of oneself: Perfectionist value perfects one's own human nature, which has to be internal to oneself. However, the main, general point that Sumner seems to be trying to bring out here is this: Prudential value is *personal* in a way that perfectionist value isn't. And this main, general point seems right. (Later in this section I will affirm that the perfectionist viewpoint is third-personal in a way that differs from the prudential viewpoint. Thus I do think that there is a sense in which the perfectionist viewpoint can properly be said to be *external*. But, that said, it must be emphasized that perfectionist value perfects one's own human nature, which, to repeat, has to be internal to oneself. Sumner neglects to mention this point, and that is a serious oversight.)

In Thomas Hurka's book *Perfectionism* Hurka makes it clear that his account of perfectionist value is not an account of well-being. Hurka states:

But the [perfectionist] ideal is not about what is "good for" humans in a more common sense...In this more common sense, "good for" is tied to the concepts of well-being or welfare and interests: Something is "good for" a person if it increases his well-being or furthers his interests. Well-being itself is often characterized subjectively, in terms of actual or hypothetical desires. Given this subjective characterization, perfectionism cannot concern well-being. Its ideal

cannot define the “good for” a human because the ideal is one he ought to pursue regardless of his desires. In my view, perfectionism should never be expressed in terms of well-being. It gives an account of the good human life, or of what is good in a human, but not of what is “good for” a human in the sense tied to well-being.<sup>111</sup>

Thus Hurka, like Sumner, maintains that there is a split between prudential value for humans (or human well-being) and perfectionist value for humans.

Further, Nicholas Wolterstorff has recently said: “Proper functioning radically under-determines well-being. One cannot read the contents of well-being off our proper functioning.”<sup>112</sup> Here Wolterstorff seems to be claiming that perfectionist value (or proper/excellent functioning) *under-determines* well-being. And this seems to take us back to one of the views that I referenced in section 16. Here I have in mind the view that, while all prudential goods are perfectionist goods, not all perfectionist goods are prudential goods.

I agree with Sumner, Hurka, Wolterstorff, and all others who hold that prudential value and perfectionist value are not identical. The truth conditions for these two things are not the same. In particular, prudential value is (in my estimation) a *desire-dependent* sort of value, whereas perfectionist value is a *desire-independent* sort of value. This naturally flows from the fact that prudential value is *personal* in a way that perfectionist value isn't. Still, a point of clarification is in order here. While I deny that prudential value and perfectionist value are identical, there is a clear sense in which my own view is very close to the view of those perfectionist OL theorists who take the identity route. This point can be brought out as follows. Wolterstorff claims that proper functioning *radically* under-determines well-being, and in saying this he seems to mean that many things that are good for us in perfectionist terms are not good for us in prudential terms. I

would reject this claim. At least if we are bracketing moral virtue, I would go so far as to claim that *every* state of affairs that is intrinsically good for me in the perfectionist sense is *also* intrinsically good for me in the prudential sense. More generally, it is fair to say that on my view the vast majority of us are made better off in prudential terms by the exact same states of affairs that make us better off in perfectionist terms – or, at any rate, this holds at least if we are bracketing moral virtue. Perfectionist OL theorists who take the identity route claim that prudential value for humans and perfectionist value for humans are one and the same thing. This claim is, I believe, false. But, from my point of view, it is easy to see why one would make this claim: One would make this claim because, at least if we are bracketing moral virtue, it so happens that the vast majority of us are made better off in prudential terms by the exact same states of affairs that make us better off in perfectionist terms. (I will discuss this matter again later – see section 30.)

OL theorists will obviously disagree with the claim that prudential values are personal such that, in spelling out the nature of prudential values, we need to refer to individuals' mental states (or pro-attitudes). On the view of OL theorists, it is enough if we simply find a way, within the confines of an objectivist framework, to account for the fact that different things are prudentially good for different people. And, of course, OL theorists can indeed account for the fact that different things are prudentially good for different people. What instantiates friendship for you will differ from what instantiates friendship for me, since, after all, we have different friends. Further, what instantiates accomplishment for a three-year-old will differ from what instantiates accomplishment for any normally functioning adult. So, yes, there clearly is a sense in which OL theories entail that prudential values are personal values (or are values that are fit specifically for

the person that they benefit). Nevertheless, when we think of prudential values as being distinctively personal, I think that we have something more than just this OL theory sort of personal fit in mind: I think that we have in mind a fit that involves someone's having a positive mental state (or pro-attitude) toward anything that is prudentially good for her.

Think of the point this way. The word 'person' typically refers to a being with mental properties. So, when we are talking about personal values, it is natural to think that we are talking about values that involve mental states. Thus far there is no serious problem for OL theorists, since, after all, the goods that OL theorists usually refer to (accomplishment, friendship, knowledge, etc.) seem, at least typically, to involve mental states. (Can you accomplish anything without employing some of your mental states? Can you engage in friendship without employing any of your mental states? The answer, in both cases, seems to be no. And the same sort of thing at least typically holds for the other basic goods.) But now, in addition to the thought that personal values involve mental states, consider the thought that a prudential value is a value that *benefits* the agent. This leads us to the claim that, when we say that prudential values are personal values, we don't just mean they are personal because they involve individuals' mental states – we must also mean that they are personal because they involve individuals' *positive* mental states. Here the word 'positive' is correlated with the word 'benefit'.

Another thing to keep in mind here is that there seems to be something essentially first-personal about prudential value that distinguishes it from, say, both moral value and perfectionist value. When we talk about what is good for A, and when (in doing so) we have prudence in mind, it seems that, at least to some extent, we have A's own viewpoint, or personal stance, on the world in mind; and, naturally, to reference A's own viewpoint



or personal stance on the world is, in some essential way, to reference A's own attitudes toward the world. Thus it does seem that to speak of what is prudentially good for A is to speak of something that A has a positive attitude toward. (While speaking of viewpoints or personal stances may seem to suggest that *belief* or *cognitive endorsement* is the appropriate attitude to seize on here, *desire* is, I think, the appropriate attitude to seize on. Seizing on belief or cognitive endorsement seems to be a bad idea – see section 33.)

The moral point of view is certainly not first-personal or attitude dependent in the way that the prudential point of view is. As I have indicated, morality seems to focus on either impartiality or constraint (or both). If impartiality is being emphasized, then, while one's own attitudes may well matter in that they matter for her own welfare, it isn't one's own welfare that is of decisive importance when one is being impartial – rather, it is overall welfare or overall goodness that is of decisive importance when one is being impartial. Alternatively, if the focus of morality is on constraint, then the point is that morality has categorical force in that it lays down certain rules that agents are required to follow, irrespective of their own attitudes. Thus, regardless of whether the focus is on impartiality or constraint, it seems clear that the moral realm isn't essentially first-personal or attitude dependent in the way that prudential realm seems to be.

Now turn to perfectionist value. As I see it, the perfectionist viewpoint is strictly third-personal (in a sense to be explained). One looks down on the world, taking into account what one sees, and one asks how it is that given kinds of things function properly or well. How do tulips function properly or well? How do golden retrievers function properly or well? How do humans function properly or well? Once you have answered this question about proper/excellent functioning for the kind of being (whatever it may

be) that is relevant to your inquiry, you can apply your answer to the individual that is relevant to your inquiry. So, if you are considering yourself, and if you are taking up the perfectionist point of view, you first try to determine (from a third-person perspective) what it takes for humans to function properly or well. Then, after that, you apply this answer to yourself, taking into account your own physiology, circumstances, abilities, personality, etc. All the while, though, you remain in the third-person mode. The question you are asking is what it takes for *any* human who has this sort of physiology, who is in these sorts of circumstances, who has these sorts of abilities, who has this sort of personality, etc., to flourish. Yes, the question here applies to you specifically, but the point is that it could be asked of *any* human that fits this description. In that sense, then, the question is general, not personal. Moreover, the question doesn't directly concern your own present attitudes – that is, although the question certainly takes into account specific things about you (your physiology, your abilities, etc.), it is nevertheless asked in such a way that it abstracts from, or brackets, your own present attitudes.

Whereas the perfectionist point of view is strictly third-personal, the prudential point of view is at least in part a first-person, agential point of view that gives direct importance or weight to your own present attitudes. This does not mean that the prudential point of view is strictly (or purely) first-personal. I am willing to say that, on my view of well-being, prudential value is partly a function of perfectionist value. And, if I do say this, then my view of well-being entails that, although the prudential point of view is in part a first-person, agential point of view, the prudential point of view is also in part a third-person point of view – that is to say, the prudential point of view incorporates within itself a third-person point of view (namely, the perfectionist point of view) as one

of its moments. The idea here would be that, when you take up the prudential viewpoint, you do ask the perfectionist question of what it takes for *any* human who has this sort of physiology, who is in these sorts of circumstances, etc., to function properly/well. But, and this is the key point here, this would just be one part of the story. For, whenever you are taking up the prudential point of view, you have to ask yourself the personal question of what is intrinsically good for *you*, or of what benefits *you*. In this way the prudential viewpoint is essentially personal – it requires you to enter into the first-person, agential mode and to give your own present attitudes direct importance. In sum: That someone’s own present attitudes are directly important – and, indeed, that they enter into the truth conditions of what makes something (anything) be an aspect of her welfare – is, I believe, an essential feature of the prudential realm.

To those OL theorists who are unmoved by these remarks, I would pose the following question: How are you going to account for the personal nature of prudential values? Then, once the OL theorist works out his or her answer to this question, I would attack the details of the answer. Here, in particular, I would highlight the clarity and simplicity of my own account of how it is that prudential values are personal (prudential values are personal values, I say, because they hook into an individual’s own actual desire-set), and I would challenge the OL theorist to come up with an account that is clearer and simpler – and, more generally, better – than my own account.

To sum up much of what I have said in this section: Prudential internalism doesn’t concern normativity in general/unqualified normativity, or morality, or perfectionist value. It only concerns prudential value. This matters, for, once the target of prudential internalism has been properly delimited, it should be clear that prudential internalism is

very plausible. Think of the matter this way. The realm of prudence is the realm of self-interest. It is, then, a realm of strong partiality. Moreover, and relatedly, human well-being is an especially *personal* sort of value. In this sense human well-being (or, in other words, prudential value for humans) differs from both moral value and perfectionist value. And, given these points, it stands to reason that prudential internalism is true – that is, it stands to reason that what is prudentially good for A must positively hook into A’s own subjective motivational set.

I will fill this argument out further in section 20. But, before we leave section 19 behind us, I want to do one more thing. I want to bring up one of the common ways of arguing for prudential internalism; and, in particular, I want to make it clear that I reject this common way of arguing for prudential internalism.

One common way of arguing for prudential internalism starts out by simply assuming the truth of subjectivism about value in general. Subjectivism about value in general asserts that all values are/must be (somehow) directly and entirely constituted by the mental states of individuals. Rosati has discussed the argument for prudential internalism that starts out by simply assuming the truth of subjectivism about value in general.<sup>113</sup> She states:

A second argument for internalism about a person’s good begins with an argument for subjectivism, according to which value exists only in virtue of subjectivity. R. B. Perry expresses this idea in *General Theory of Value*: “A certain positive plausibility is given to this hypothesis by the fact that in order to create values where they did not exist before it seems to be sufficient to introduce an interest. The silence of the desert is without value, until some wanderer finds it lonely and terrifying; the cataract, until some human sensibility finds it sublime, or until it is harnessed to satisfy human needs.” Introduce into the world creatures who are affected by and react to their world, Perry tells us, and you introduce value as well. Indeed, he suggests, introducing such creatures is sufficient for the introduction of value...It is a natural step from this to the thought that value itself must be a complex motivational property. If value can exist only if there are

creatures who can be affected by and react to their world, then value, and more specifically, goodness for a person, must be a motivational property.<sup>114</sup>

I am not attracted to this argument for prudential internalism, for I reject subjectivism about value in general. Subjectivists may protest: ‘But don’t you agree that, if there were a world without any minded individuals, then this world wouldn’t contain any value? And doesn’t this show that subjectivism about value in general is true? So, in arguing for prudential internalism, shouldn’t we start out by assuming the truth of subjectivism about value in general?’ To this I have two things to say.

(1) Kraut has argued that plants have a welfare, and also that what is intrinsically good for plants is obviously not directly constituted by mental states.<sup>115</sup> I don’t know whether plants really have a welfare – or whether there really is such a thing as prudential value for plants.<sup>116</sup> But, regardless of whether plants have a welfare or not, it seems that there are things that are intrinsically good for plants in the *perfectionist* sense. After all, it is completely natural to speak of things that are good or bad for plants, and it seems right to think that this way of speaking is genuinely referential. (For instance, if I stomp on tulips, this is bad for these tulips.) Supposing, then, that there is a world that contains plants but no minded individuals, it seems to me that there is value in this world, for it seems to me that this world contains perfectionist value for plants.

(2) Even if we are just talking about humans and their welfare, it seems implausible to hold that mental states are doing all of the direct constitutive work – it seems implausible to hold that human welfare *fully reduces* to mental states. In order to hold this view, one would need to accept a DF theory, or a version of welfare hedonism, or Sumner’s welfare theory, or something close to one of these three sorts of subjectivist views of well-being. But none of these subjectivist views strike me as being plausible.

(Incidentally, I don't mean for this to count as a thorough argument – I am just trying to make my own position on prudential internalism clear.)

To sum up: In arguing for prudential internalism, it seems that the argument shouldn't start out by simply assuming the truth of subjectivism about value in general. It seems that only a different line of argument will work.

## **20. Prudential internalism: interpersonal justification, intrapersonal deliberation**

Many philosophers with strong normative-objectivist tendencies are naturally averse to the thesis of prudential internalism, at least in any strong form. The main aim of the previous section was to soften these philosophers up a bit. Indeed, the main aim was to get these philosophers to see the following: It is sensible to be averse to existence internalism as applied to, say, the realm of unqualified normativity, or as applied to the realm of perfectionist value – but it isn't sensible to be averse to prudential internalism, taken just in itself. In this section I am going to argue for prudential internalism in what is, perhaps, a more direct way. Here, in particular, I will focus on our interpersonal justificatory practices concerning well-being, and also on the intrapersonal prudential decision-making process.<sup>117</sup>

OL theorists hold that the states of affairs that are prudentially good for someone are so whether she has any sort of pro-attitude toward them or not – and, indeed, even if she is thoroughly averse to them. This is a natural consequence of the way that OL theorists lay out the truth conditions for something's (anything's) counting as an aspect of A's welfare. According to OL theories, the truth conditions for something's (anything's) counting as an aspect of A's welfare don't make any reference to A's desires (or, more

generally, to A's pro-attitudes). For example, if X instantiates aesthetic experience for A, then it follows on OL theories that X is an aspect of A's welfare, and that X is so whether A has any sort of pro-attitude toward X or not (and, indeed, even if A is thoroughly averse to X). All of this seems to cut against our everyday intuitions. For, while there may be some exceptions, the general rule seems to be that people don't tend to think of the aspects of their own welfare as things that they could be thoroughly averse to.

Here OL theorists may reply that everyday intuition is actually on their side; and here they may point to the fact that in ordinary language we often say things like 'this is good for you whether you want it or not' and 'I didn't want to do it, but I must concede that doing it was good for me'. I grant that we often say these sorts of things, but I think that, when we say these sorts of things, we don't have *intrinsic* prudential value in mind – we only have *instrumental* prudential value in mind. Suppose, for instance, that Shakira says: 'I know that taking the medicine is good for me, but I don't want to take it'. Taking the medicine is only *instrumentally* prudentially good for Shakira here. The healing that taking the medicine brings is what is *intrinsically* prudentially good for Shakira here. Thus it is wrong to think that, in not wanting to take the medicine, Shakira lacks the desire for some aspect of her welfare. Another example: Suppose that Brad takes the train from DC to New York every other weekend to see his girlfriend, Sharon. Brad may say: 'I know that making the trip every other weekend is good for me, but I hate those long train rides – I really don't want to make the trip.' Making the trip is only *instrumentally* prudentially good for Brad here. It is, say, engaging in friendship with Sharon that is *intrinsically* prudentially good for Brad here. So it is wrong to think that, in not wanting to take the train ride, Brad lacks the desire for some aspect of his welfare.

*Our interpersonal justificatory practices concerning well-being:* Consider Shakira again. If we are trying to convince her that she should take the medicine because doing so will promote her well-being, then it seems that we will say: ‘Shakira, we know that you don’t want to take the medicine. That is natural – it tastes bad. But, if you don’t take the medicine, then you won’t heal. And you *want* to heal, don’t you?’ It is not that, in saying this to Shakira, we are assuming that her desire for healing *by itself* directly makes her healing be intrinsically prudentially good for her. Rather, it is that, in saying this to Shakira, we are assuming that her healing cannot be an aspect of her welfare unless her healing positively hooks into her desire-set. Moreover, it would be bizarre if we were to say: ‘Shakira, you should take the medicine because by doing so you’ll get something – namely, some healing – that will add to your well-being. Oh, and by the way, it is simply *irrelevant* whether you want to heal. After all, your healing will add to your well-being whether you want to heal or not.’ In justifying to someone that something is prudentially good for her, I don’t think that we would discount her desires in this way.

The same holds with respect to Brad. If we are trying to convince him to make the trip to New York because doing so will promote his well-being, then we will say something like this: ‘Brad, we know that you don’t want to make the trip to New York. But these are the circumstances of your relationship with Sharon. You don’t live in the same city, and so you have to trade off making the trip. If you don’t make the trip every other weekend, then your relationship probably won’t go well. And you *want* your relationship to go well, don’t you?’ Again, it’s not that we are assuming that Brad’s desiring *that his relationship with Sharon go well* is something that *by itself* directly makes this state of affairs be an aspect of his welfare. Rather, it is that we are assuming



that this state of affairs can't be an aspect of Brad's welfare unless it positively hooks into his desire-set. Also, it would be weird if we were to say: 'Brad, you should take the trip to New York, for doing so will help you get something – namely, a healthy relationship with Sharon – that will add to your welfare. Oh, and by the way, it is simply *irrelevant* whether you want to have a healthy relationship with Sharon. After all, your having a healthy relationship with her is part of your welfare whether you want this or not.'

Again, in justifying to someone that something is prudentially good for him, I don't think that we would discount his desires in this way.

My general point here is this. When we are justifying to others that they should do such-and-such because by doing such-and-such they will promote their own well-being, we typically do appeal (at least implicitly) to their desires. Indeed, the justification enterprise seems (at least implicitly) to have *two* elements to it, both of which are essential: There seems to be an appeal to the content of the state of affairs that is thought to be an aspect of the agent's welfare *and* a linking of this state of affairs to the agent's desire-set. Admittedly, with some agents – for instance, little babies – we don't try to do any persuading that links the putative aspect of welfare to the agent's desires. Still, the general rule here is that, when we are trying to persuade someone that such-and-such is intrinsically prudentially good for her, we do try (at least implicitly) to bring out a direct connection between what is intrinsically prudentially good for her and her desire-set. So, when we engage in our interpersonal justificatory practices concerning well-being, we seem to assume that prudential internalism is true (and is so in a relatively strong form).

*Intrapersonal prudential deliberation:* Another point to consider here is this. It would be peculiar if someone were to think to himself: 'Oh, I should definitely do this,

for by doing this I'll get X, which adds to my well-being. Of course, the question of whether X positively hooks into my desire-set is simply *irrelevant*. Indeed, my desires don't matter here – not at all.' Individuals don't discount their desires in this way when they are thinking about their own well-being. On the contrary, in the case of justifying to oneself that X is part of one's well-being, the justification process at least implicitly takes into account *not only* the content of X, but *also* whether X hooks into one's desire-set.

An example may help. Suppose that Lilly is choosing between two graduate schools for philosophy: Q and R. Further, suppose that, while talking out her decision with one of her friends – let's call the friend Olivia – Lilly says: 'You know, Olivia, there are various factors that matter here. Q is only two hours from home, and so going to Q would allow me to see my family and friends from home pretty often. Also, Q is a better department overall than R, at least according to the Leiter report. I should take that into account. And, you know, the campus at Q is very pretty. The beauty of the campus might not be the most important thing to consider when you are choosing a grad school. But let's be honest: Everybody wants to be in an aesthetically pleasing atmosphere. Of course, R is notably stronger than Q in epistemology, and I do think that I want to focus on epistemology. And there is also the issue of money to consider: The stipend at R is much better than at Q, while the cost of living would be about the same at either place. I don't want to have to worry too much about whether I can pay my rent, can pay my bills, can pay for groceries, can pay for transportation, etc. All and all, I don't know what to do: I want to go to Q for various reasons, but I also want to go to R for various reasons.'

When we engage in intrapersonal prudential decision-making, it seems that we typically do what Lilly is doing here: We go back and forth between various features of

the options that are open to us (on the one hand) and our own desires (on the other hand). Or, put more fully, what we typically do is this. We focus on various things that are valuable for us in a desire-independent way: things like friendship, accomplishment, knowledge, pleasure, health, and aesthetic experience. Further, we consider these desire-independent goods in relation to our own circumstances, abilities, etc., and we consider the ways in which – and also the extent(s) to which – the options open to us instantiate these desire-independent goods for us. Yet, and this is the key point here, this is not *all* that we consider when we are making prudential decisions. For we also consider our own *desires* – that is to say, we also consider what we *want*. In thinking about Q and R, for example, it is perfectly natural for Lilly to consider her own desires (as she is here doing). Moreover, it would be very strange if Lilly were simply to ignore her own desires.

In order to bring this point out more clearly, let's imagine that, after listening to Lilly's thoughts on Q and R, Olivia looks at Lilly and says: 'Lilly, I think that there is a major problem with the way that you're approaching this grad school decision. You keep thinking about what you *want*. That is a bad move. Here is my advice: Forget about your own desires; they just aren't of any direct relevance here. Speaking generally, people just shouldn't think about what they want when they are making really important prudential decisions. Indeed, when people are making prudential decisions about where to go to grad school for philosophy, or about who to date seriously, or whatever, the best move for them is to ignore their own desires.' This is, I think, weird and bad advice. Nobody does or should discount her own desires in this way when she is engaging in prudential decision-making. But notice that, if OL theories were true, then Olivia's advice would not be so weird or bad. After all, OL theories entail that Lilly's own desires don't do any

of the direct constitutive work vis-à-vis Lilly's own welfare. Thus, if OL theories were true, then Olivia would actually be quite correct in telling Lilly that her own (that is, Lilly's own) desires aren't of any direct relevance in this context.

What might OL theorists say in response to my comments about interpersonal justification vis-à-vis well-being and intrapersonal prudential deliberation? Here two sets of points are in order.

One: OL theorists may say this. 'Whenever you are deliberating, your desires have to be given direct importance. This isn't just because, at least typically, an act can't get off the ground without the help of some desire – this is also because, at least typically, deliberation starts with a consideration of desired ends. With respect to interpersonal justification vis-à-vis well-being and intrapersonal prudential deliberation, it is easy to jump from the thought that someone's desires are crucial if she is engaging in prudential deliberation to the thought that someone's desires must enter into the truth conditions for what makes something be prudentially good for her. But this jump is illegitimate.'

I agree that the role that desires play in prudential deliberation is one thing, while the role that desires play in the truth conditions for something's counting as an aspect of welfare is something else. But, that said, I don't think that my comments in this section betray any confusion in this regard. My point has been that, when we justify to someone that something is included within her well-being, we typically try (at least implicitly) to show her that the thing in question hooks into her desire-set; and, when we do this, it seems that we are assuming that the thing in question can't be an aspect of her welfare unless it hooks into her desire-set. Also, my point has been that, when someone is trying to figure out what is prudentially good for herself, she will give some weight to her own

present desires – not just because deliberation always involves desires, but also because she will be assuming (at least tacitly) that something must hook into her desire-set in order for it to be prudentially good for her. To those OL theorists who remain unmoved by my remarks, I would go back to something that I noted in the last section. I would stress that prudential values are personal (in a way that moral values and perfectionist values aren't), and I would say that OL theorists must account for this. Then I would highlight the clarity and simplicity of my own account of how it is that prudential values are personal (prudential values are personal values, I say, because they hook into one's own actual desire-set), and I would challenge the OL theorist to come up with an account that is clearer and simpler – and, more generally, better – than my own account.

Two: Murphy is an OL theorist who may be able to handle much of what I have said about interpersonal justification vis-à-vis well-being and intrapersonal prudential deliberation. I say this because Murphy puts inner peace – which is the good of having no unsatisfied desires – on his list of basic goods.<sup>118</sup> On Murphy's view, then, although it is true that A's desires don't do any of the direct constitutive work vis-à-vis A's welfare, there nevertheless is a very clear sense in which A should take her own desires into account when she is engaging in prudential deliberation. However, Murphy's claim that inner peace counts as a basic good is controversial: No other OL theorists (explicitly) accept this claim of Murphy's. Also, it should be said that inner peace doesn't look like any of the other basic goods. Health, knowledge, friendship, aesthetic experience, etc., are all such that, when one thinks of them, one immediately thinks of something that is intrinsically good for oneself. But, when one thinks of inner peace, it isn't clear that one thinks of something that is intrinsically good for oneself. My general point here, then, is

this. If an OL theorist appeals to inner peace (or something like it) as a way of giving importance or weight to an individual's desires, then, while this is good in that it displays a willingness to admit that desires matter a great deal for human well-being, there nonetheless seems to be something amiss with (or not quite right about) this particular way of trying to accommodate the importance of desires for human well-being.

To be clear about my general point here in section 20: Human welfare *does* largely depend on desire-independent goods such as knowledge, friendship, and health. OL theories have this right. My general point here in section 20 is just that there is something significantly *incomplete* about the OL theory picture. In particular, one's own desires can't be discounted in the way that OL theories discount them.

## **21. OL theories and prudential internalism: why the hypothetical route fails**

If what I have said in section 20 is on target, then we must accept a relatively strong form of prudential internalism – in particular, we must accept a form of prudential internalism that links the aspects of A's welfare to A's desire-set. But this means that OL theories are in trouble. After all, OL theories don't build any reference to A's desires into the truth conditions for something's (anything's) counting as an aspect of A's welfare.

What, then, can OL theorists do here? At this point it is common for OL theorists to move to a version of *hypothetical* prudential internalism – that is, to a version of prudential internalism that links the aspects of A's welfare not to A's actual desires, but rather to the desires that A would have, given that certain conditions were to obtain. The OL theorist claim here is always roughly this: Given that X is an aspect of A's welfare, it must be true that A would desire X if A were properly to appreciate the nature of X (and

here OL theorists will fill out ‘properly appreciate the nature of X’ in a way that accords with an OL theory account of human welfare). This OL theorist claim derives from Griffin, who says that prudential values “are, on the most plausible account of their link to desire, what one would want if one properly appreciated the object of desire.”<sup>119</sup>

In this section I will argue that the OL theory move to hypothetical prudential internalism doesn’t work. It doesn’t work because it doesn’t do much (if anything at all) to ensure that an individual won’t be alienated from aspects of her own welfare.

Here we can start by considering straightforward prudential internalism, which says that, if X is an aspect of A’s welfare, then it must be true that A desires X or has some other sort of pro-attitude toward X. Actual DF theorists accept straightforward prudential internalism. However, speaking generally, very few philosophers are willing to accept this thesis. The natural worry here is that there may well be various things that are part of A’s welfare, and yet that A does not have any pro-attitude toward – either because A doesn’t know about them or for some other reason. Rosati puts the rejection of straightforward prudential internalism like this:

A person’s good [that is, a person’s well-being] will, of course, include many things that she could antecedently desire. But it will also include many things that she did not or could not antecedently desire, as well as things that she can acquire only by accident, or at least not by direct effort. A person can fall in love, for instance, with someone that he met by chance, or benefit from the kindness of strangers, or delight in a spontaneous adventure. Our good is perhaps achieved as much through serendipity as through rational endeavor.<sup>120</sup>

Suppose that we go ahead and reject straightforward prudential internalism. What, then, is the next move?

At this point many philosophers think: ‘Given that X is an aspect of A’s welfare, there has to be some fairly tight connection between A’s pro-attitudes and X. Yet we’ve

already rejected straightforward prudential internalism. So we've already rejected the claim that A must have some *actual* pro-attitude toward X. There is, then, just one option left: We must claim that X must hook into A's *hypothetical* pro-attitudes (or into the pro-attitudes that A would have, given that certain conditions were to obtain).'

On my view we should, if at all possible, simply avoid the move to hypothetical prudential internalism.<sup>121</sup> But let's leave my own view to the side here. For the moment, let's just focus on the question of whether there are there any acceptable versions of hypothetical prudential internalism. The main intuition behind prudential internalism is this. It can't be true that X is *both* an aspect of A's welfare *and* thoroughly alienating to A. The most natural way to formulate prudential internalism is in terms of A's actual pro-attitudes: We should simply say that, in order for X to be an aspect of A's welfare, X must positively hook into A as she actually is. Once we abandon actual prudential internalism and instead frame prudential internalism in terms of some hypothetical pro-attitude of A's, this seems inevitably to open up the space for cases where, although hypothetical A is positively linked to X (where X is an aspect of actual A's well-being), actual A isn't positively linked to X in any way at all. Yet, once we are working with a version of prudential internalism that allows for actual A to be alienated from aspects of her own well-being in this way, it is hard to understand why we are still trying to advance any version of prudential internalism at all. Rosati has come up with a certain version of hypothetical prudential internalism that aims to do away with worries concerning alienation. It will help, I think, if we spend some time examining Rosati's version of hypothetical prudential internalism. For, once we have done this, we will be in a good position to comment on the plausibility of the OL theory version of hypothetical



prudential internalism (which, to repeat, says that X can't be an aspect of A's welfare unless A would desire X if A were properly to appreciate the nature of X).

Rosati eventually winds up with what she refers to as *two-tier internalism*. But her reflection begins with what she refers to as simple internalism. She says:

[A]s long as we can care about or like or be glad of something once we acquire it or experience it, this seems enough to satisfy the intuition behind internalism. Internalism about a person's good should thus not be characterized in relation to motivation narrowly construed. Rather, it should be characterized in relation to motivation in the broad sense. In this sense, to motivate is to prompt or elicit a pro-attitude – such as desiring, liking, being glad of, caring about, and so on – an attitude which may or may not be a motivation to action. To say that something must motivate, in the broad sense, to be part of a person's good, is to say that it must be something that can, in a positive way, matter to her or be an object of her concern. For simplicity, I shall from here on speak in terms of caring about a thing. This locution is not intended to designate the strong sense of caring about a thing that contrasts with merely liking or wanting it. Instead, 'caring about' should be understood to stand for having one or another of those pro-attitudes that we commonly have toward things plausibly regarded as a part of our good. Anything that a person is capable of caring about in this sense is a possible object of her concern. Let's understand simple internalism, then, as follows: something X can be good for a person A only if A is capable of caring about X. X, to be good for A, must be a possible object of her concern.<sup>122</sup>

Rosati rejects simple internalism, saying that it is "clearly too weak: it counts as possible goods for a person things that would violate the intuition that inspires internalism – the intuition that a person's good must suit her, that it cannot be something alien to her."<sup>123</sup>

As indicated earlier in this section, though, Rosati also rejects straightforward prudential internalism (which says that X can't be an aspect of A's welfare unless A has some actual pro-attitude toward X). This leaves Rosati's view somewhere in-between straightforward prudential internalism (which she sees as being too strong) and simple internalism (which she sees as being too weak). Rosati puts the matter as follows:

If internalism is to remain faithful to the intuition that inspires it, it must be formulated so as to avoid two extremes. On the one hand, it must not hold that something can be good for a person only if the actual, unaltered person can care

about it, for this ties a person's good too closely to her present condition, however defective. On the other hand, although it must allow that, in some sense, anything she is capable of caring about might be good for her, it must not include as possible goods for her what she would care about only under alienated conditions. Internalism must treat a person's capacity to care counterfactually, while constraining counterfactual conditions so that they permit as possible goods for a person only what can recognizably fit or suit her.<sup>124</sup>

Rosati eventually settles on two-tier internalism. She says:

The stronger form of internalism that we finally arrive at tells us that something X can be good for a person A only if two conditions are met: (1) Were A under conditions C and contemplating the circumstances of her actual self as someone about to assume her actual self's position, A would care about X for her actual self; (2) conditions C are such that the facts about what A would care about for her actual self while under C are something A would care about under ordinary optimal conditions.<sup>125</sup>

What, though, does this mean? Here we can explain the first tier of Rosati's thesis (or tier #1) and then move onto the second tier of Rosati's thesis (or tier #2).

*Tier #1:* Rosati is here making it clear that she rejects straightforward prudential internalism. On Rosati's view it is certainly possible for various things to be part of A's welfare even though A has no actual pro-attitude toward them – for instance, because A doesn't know about them, or because A isn't thinking clearly with respect to them. The aim of tier #1, then, is to move A out of the conditions that she is actually in at the present moment and into conditions that are somehow more appropriate (or more ideal). Rosati leaves it open as to what these more appropriate (or more ideal) conditions (that is, conditions C) might be. Indeed, Rosati says that her two-tier internalism “says nothing about what the correct counterfactual conditions might be”, and in saying this Rosati means that her two-tier internalism says nothing about what conditions C are.<sup>126</sup> (Of course, the natural thought is that conditions C will inevitably be filled out as being, perhaps among other things, conditions wherein A is provided with much more non-

evaluative information than she has in her present, actual state.) Even though Rosati's two-tier internalism is silent/agnostic about what conditions C might be, Rosati's two-tier internalism does place an important constraint on conditions C. This important constraint on conditions C is captured by tier #2 of Rosati's two-tier internalism.

*Tier #2:* However conditions C are filled out, Rosati thinks that they must be filled out in a way that isn't alienating to A. Rosati puts the point as follows:

How are we to constrain what shall count as "appropriate" conditions [that is, conditions C]?...[W]e must insure that these conditions do not themselves strike us as alien. Yet how conditions may now sit with a person need not settle their appropriateness any more than whether she can now care about something settles whether it is good for her. We need a way to rule out alienated conditions, without allowing the appropriateness of conditions to depend upon how they now strike a person, whatever her present state might be...We can do so by requiring that a person regard counterfactual conditions as appropriate not in her present state, but under ordinary optimal conditions. These would include that a person not be sleeping, drugged, or hypnotized, that she be thinking calmly and rationally, and that she not be overlooking any readily available information. This list of conditions is not meant to be exhaustive. By 'ordinary optimal conditions' I mean whatever normally attainable conditions are optimal for reflecting on questions about what to care about self-interestedly. Thus, we might say, counterfactual conditions [that is, conditions C] are appropriate only if a person would so regard them under ordinary optimal conditions.<sup>127</sup>

Thus, whatever conditions C turn out to be (again, Rosati intentionally leaves this open), it must be true, according to Rosati's two-tier internalism, that conditions C are such that A herself would regard them as being appropriate, provided that A herself were in ordinary optimal conditions. (Or, since Rosati actually goes on to revise this last point slightly, let's just say that this last point is at least roughly correct.<sup>128</sup>) There is a question that arises here: Why doesn't Rosati just go with a version of hypothetical prudential internalism that says that X can't be an aspect of A's welfare unless A would have a pro-attitude toward X, provided that A were in ordinary optimal conditions? With respect to this question, Rosati says: "Ordinary optimal conditions are insufficiently ideal to be the

test of what is good for a person. Determining what it makes sense for someone to care about may require far more information, for instance, than is readily available.”<sup>129</sup>

Slightly oversimplifying, we can summarize two-tier internalism as follows: (1) X can't be an aspect of A's welfare unless A would have a pro-attitude toward X if A were placed in idealized conditions C, and (2) idealized conditions C – however they are to be filled out – must be filled out in a way that A herself would endorse if she were thinking about the matter in a calm and attentive way. Or, as Rosati summarizes her own view:

[I]f an account of a person's good is to satisfy the intuition that drives internalism, it must effect a double link to motivation. The first link provides for a motivational connection between an individual and her good under counterfactual conditions, and thus at a further remove from her actual self. This link captures the thought that while a person's good might include things that she cannot in her present state care about, it must include only what she is capable of caring about. It avoids the alienation of treating as good for a person things that cannot matter to her, while not connecting a person's good too closely to her actual concerns. The second link effects a motivational connection between an individual and information about her counterfactual attitudes. It provides for a closer connection to the actual individual, requiring that appropriate conditions connect with what a person would care about when contemplating her good under ordinary optimal conditions. This requirement further captures the intuition that a person's good must connect with what she would find compelling, at least if she were rational and aware. The second link thus prevents the alienation of treating as possible goods for a person things she would come to care about only under alienated conditions, conditions irrelevant to her inquiry about her good. And it achieves this without implausibly requiring that a person's counterfactual concerns, or even information about her counterfactual concerns, now motivate her no matter what her current condition might be.<sup>130</sup>

Now that we have on hand a sense of how two-tier internalism is meant to work, we might go ahead and briefly consider two tier-internalism in relation to alienation.

With respect to this matter, suppose that there is a version of one-tier hypothetical prudential internalism that is filled out roughly as follows: X can't be an aspect of A's welfare unless A would desire X after having just been fully and vividly informed with respect to value-free information. Here things may turn out such that hypothetical A

desires X (where X is an aspect of A's welfare), while actual A doesn't desire X and, more generally, has no pro-attitudes toward X. Further, on one-tier internalism the story ends right here. But this is too alienating: There is no positive connection between actual A and X here; and, in addition, actual A has had no say whatsoever in determining the counterfactual conditions that are being relied on here. On Rosati's two-tier internalism this alienation problem is to some extent mitigated, for the counterfactual conditions specified in the first tier must be endorsed by A – if they aren't endorsed by A, then they will be abandoned in favor of a set of counterfactual conditions that A does endorse (or, to be more precise, that A would endorse, provided that A were in ordinary optimal conditions).

I am going to leave open the question of whether Rosati's two-tier internalism does an adequate job with respect to the issue of alienation. There is no need for me to make a final judgment on this matter, for I am here simply concerned to show that the OL theory version of hypothetical prudential internalism fails.<sup>131</sup> For our purposes here in section 21, the really important point that Rosati's discussion of two-tier internalism helps to bring out is this: In order for the OL theory version of hypothetical prudential internalism (which says that X can't be an aspect of A's welfare unless A would desire X if A were properly to appreciate the nature of X) to have any shot at adequately handling alienation, it needs to incorporate a second tier (or something like a second tier).

Think of the matter this way. The OL theory version of hypothetical prudential internalism ensures that, if X is an aspect of actual A's welfare, then *hypothetical* A (who is a proper appreciator of prudential values) has some sort of pro-attitude toward X. Yet this doesn't do much (if anything at all) to remove worries concerning *actual* A's being

alienated from aspects of her own welfare. What, then, can OL theorists do to mitigate this alienation problem? The natural thought is that they might take up Rosati's idea of adding a second-tier – one that says that the hypothetical conditions specified in the first tier must be such that A would endorse them if A were in ordinary optimal conditions.

However, OL theorists can't add a second tier (or anything like a second tier) to their version of hypothetical prudential internalism. To add a second tier (or something like a second tier) is, roughly put, to give A the option of endorsing or not endorsing the hypothetical conditions specified in the first tier of the OL theory version of hypothetical prudential internalism. But OL theorists obviously can't give A this option. After all, if OL theorists give A this option, and if A doesn't endorse the hypothetical conditions specified in the first tier of the OL theory version of hypothetical prudential internalism, then OL theorists are left without any version of prudential internalism. (OL theorists don't have the luxury of bringing up various sets of hypothetical conditions that might be relied on in the first tier, all with the aim of finding one that actual A endorses. For OL theorists, there is only one set of hypothetical conditions that will do: The hypothetical conditions must be ones wherein A is a proper appreciator of prudential values – and where 'proper appreciator of prudential values' is being understood in OL theory terms.)

To clarify: When OL theorists advance their version of hypothetical prudential internalism, they are thinking as follows. 'Our version of prudential internalism – which says that X can't be an aspect of A's welfare unless A would desire X if A were properly to appreciate X's nature – is acceptable. Yes, we have moved to a counterfactual here. But that isn't a problem. After all, almost all non-OL theories rely on counterfactual versions of prudential internalism. What is the difference between the non-OL theory

versions of hypothetical prudential internalism and our version of hypothetical prudential internalism? You might say that the difference is that the non-OL theory versions are stronger and that they don't allow for alienation. But in fact they do allow for some rather bad alienation.' At this point, however, the non-OL theorist who accepts Rosati's two-tier internalism can reply that the alienation problem can be mitigated by the addition of the second tier; and this non-OL theorist can add that OL theorists don't have recourse to the addition of the second tier. And this is true: OL theorists don't have recourse to the addition of the second tier. The conclusion to draw here, then, is this: OL theories have no hope of providing us with a version of prudential internalism that is sufficiently non-alienating. (I am not saying that two-tier internalism is sufficiently non-alienating – I am only saying that, unlike the OL theory version of hypothetical prudential internalism, two-tier internalism at least has a fighting chance of being sufficiently non-alienating.)

Let's now close things out for Chapter 2. OL theories aren't far from the truth about human welfare. I say this because human welfare obviously does largely depend on desire-independent, relational goods such as friendship, accomplishment, and health. (Note that it is a plain fact that people think about their well-being directly in terms of friendship, health, etc.) However, OL theories deny that someone's well-being directly depends on her own desires (or, more generally, on her own pro-attitudes); and this denial is implausible. (Note that it is also a plain fact that people think about their well-being directly in terms of their own desires.) In Chapter 3 we will focus on the no priority theory, which is distinctive and compelling because it pays due respect *both* to the natural intuition that A's welfare directly depends on what instantiates friendship, health, etc., for A *and* to the natural intuition that A's welfare directly depends on A's own desires.

## Chapter Three: The No Priority Theory

### 22. The no priority theory: comments on its value and desire conditions

The prudential problem is the problem of capturing both the *good* and the *for* parts of ‘good for’. DF theories have trouble with the *good* part of ‘good for’, and OL theories have trouble with the *for* part of ‘good for’. The no priority theory seeks to capture both the *good* and *for* parts of ‘good for’ by building a value condition and a desire condition into the truth conditions for something’s (anything’s) counting as an aspect of someone’s welfare. As indicated earlier (see section 2), the no priority theory says that X is an aspect of A’s welfare if and only if – and directly because – (1) X is, for A, an instance of one of the basic goods (that is, of health, or friendship, or knowledge, etc.) and (2) X’s obtaining satisfies some desire that A has (with the idea being that, if A does not desire X itself, then A must at least desire the basic good that X here instantiates for A).

Earlier (in section 2) I briefly discussed the no priority theory’s value and desire conditions. I will now expand upon the remarks that I made there.

*The value condition:* The no priority theory is like OL theories in that it relies on the basic goods (friendship, aesthetic experience, etc.). But there is a key difference between my conception of these goods and the OL theory conception of these goods. Like OL theorists, I conceive of these goods as intrinsic, desire-independent, and relational goods. But, unlike OL theorists, I conceive of these goods as *non-prudential* goods. If need be, I am willing to conceive of the basic goods as *perfectionist* goods. Thus we can summarize my view of the basic goods as follows: They are non-prudential, intrinsic, desire-independent, and relational goods – or, if I must label them, they are



perfectionist goods. (Later I will say more about my view of the relationship between prudential value and perfectionist value – see below, section 30.)

One thing to keep in mind here is that, just as OL theories need a list of basic goods, so too my theory needs a list of basic goods (or, perhaps better put, so too my theory needs a specific interpretation of what is on *the* list of basic goods). Different OL theorists put forth different lists of basic goods; and this gives rise to a common objection to OL theories – namely, the objection that says that, since OL theorists can't even agree among themselves about what is on *the* list of basic goods, it follows that non-OL theorists don't have much reason to take OL theories seriously. As applied to my theory, then, the objection here would be something like this: Unless and until there is an agreed upon interpretation of what is on *the* list of basic goods, we shouldn't accept the no priority theory. However, this objection seems to hold very little (if any) weight. I say this for multiple reasons. One such reason is this: Once we start looking closely at the different lists that different OL theorists advance, we can see that there is actually some substantial agreement among OL theorists when it comes to the question of what is on *the* list of basic goods.

Here let's briefly consider three different OL theory lists: Finnis's list, Griffin's list, and Alfonso Gomez-Lobo's list. Finnis's list includes: life, knowledge, play, aesthetic experience, friendship, religion, and practical reasonableness.<sup>132</sup> Griffin's list includes: accomplishment, the components of human existence (for example, autonomy and liberty), understanding, enjoyment (under which Griffin includes "the perception of beauty"), and deep personal relations.<sup>133</sup> And Gomez-Lobo's list includes: life, family, friendship, work and play, aesthetic experience, knowledge, and integrity.<sup>134</sup> No doubt

there are some differences among these three lists. But there are also some substantial similarities. Knowledge (or understanding) is on all three lists. Aesthetic experience (or, as Griffin puts it, “the perception of beauty”) is (in effect) on all three lists. Friendship (or deep personal relations) is on all three lists. Accomplishment is on Griffin’s list; and, once we read about what Finnis and Gomez-Lobo, respectively, mean by practical reasonableness and work and play, respectively, we will see that practical reasonableness (as Finnis conceives of it) includes accomplishment, while work and play (as Gomez-Lobo conceives of it) includes accomplishment.<sup>135</sup> And health is (in effect) on all three lists in that, for Griffin, health seems to fall under the components of human existence, while, for Finnis and Gomez-Lobo, respectively, health falls under life.<sup>136</sup>

The key thing to see here, then, is this. There is a certain core of basic goods – knowledge, aesthetic experience, friendship, accomplishment, and health – that are (in effect) on every (or at least almost every) OL theorist’s list. In saying this I don’t mean to downplay the fact that there are crucial disagreements among OL theories when it comes to the question of what is on *the* list of basic goods. For instance, there are crucial disagreements concerning pleasure, life, autonomy (or freedom), and moral virtue, respectively. My point here is just that, in spite of these disagreements, the fact remains that all (or almost all) OL theorists agree with one another about a certain core of basic goods. And that, I think, is enough to answer the objection that says that we should reject theories (such as OL theories and the no priority theory) that rely on an interpretation of what is on *the* list of basic goods.

*The desire condition:* The desire condition of the no priority theory says that X can’t be an aspect of A’s welfare unless X’s obtaining satisfies a desire that A has. Here

I have in mind intrinsic desires. I do not have merely instrumental desires in mind. (Again, unless I note otherwise, we should always assume that by ‘desire’ I mean ‘intrinsic desire’. Also, as I noted in section 5, it might be true that what we commonly refer to as intrinsic desires are in fact the only desires there are.) Further, in saying that X can’t be an aspect of A’s welfare unless X’s obtaining satisfies a desire that A has, I mean to be saying that X can’t be an aspect of A’s welfare unless X’s obtaining satisfies a desire that A *actually* has.

Even though the no priority theory relies on *actual* desires, the no priority theory does not rely on straightforward prudential internalism (which says that X can’t be an aspect of A’s welfare unless A actually has some sort of pro-attitude toward X). Rather, what the no priority theory relies on is a version of actual prudential internalism that says that, in order for X to be an aspect of A’s welfare, X’s obtaining must satisfy some desire that A actually has. It is important that the no priority theory does not rely on straightforward prudential internalism. Here we might go back to the case of the young academic (see sections 7 and 8). With respect to the young academic, there is a period of time during which (1) she has lost the desire to finish her book (and, more generally, has no pro-attitude toward finishing her book), during which (2) finishing her book instantiates accomplishment for her, and during which (3) finishing her book seems to be included within her welfare (as a *merely near-possible* aspect of it). Still, even though she doesn’t desire to finish her book during this period of time, she does desire accomplishment during this period of time. And, further, it is true during this time that finishing her book satisfies her desire for accomplishment. Straightforward prudential internalism is in trouble when it comes to cases such as this one. But the no priority

theory – which doesn't rely on straightforward prudential internalism – can handle this case. (Similar remarks apply to the bluegrass case and the Mediterranean case – both of which were discussed in section 8. That is to say, these are cases that pose a problem for DF theories and, more generally, for all welfare theories that rely on straightforward prudential internalism – but these cases are not a problem for the no priority theory.)

In the next section (section 23) I will lay out the main objection to the no priority theory: the missing desires objection. Then, in the five sections after that (sections 24-28), I will try to answer this objection.

### **23. The main objection to the no priority theory: the missing desires objection**

At least at first glance, the no priority theory appears to solve the prudential problem. After all, the value condition seems to guarantee the *good* part of 'good for', while the desire condition seems to guarantee the *for* part of 'good for'.

However, there is a crucial objection to the no priority theory that we should now confront. OL theorists commonly claim that DF theories cannot handle the fact that someone can fail to desire various things that look for all the world to be prudentially good for her – say, things such as accomplishment, friendship, knowledge, and health. And this same objection, suitably altered, can be – and surely will be – leveled against the no priority theory. Here we might consider a quote from Brink that I brought up earlier (see section 14). In this quote Brink is objecting to hypothetical DF theories (which he refers to as idealized desire-satisfaction views). Brink states:

If the process of idealization [or, roughly put, the process of fully and vividly informing agents] is purely formal or content-neutral, then it must remain a brute and contingent psychological fact whether suitably idealized subjects would care about things we are prepared, on reflection, to think valuable. But this is

inadequate, inasmuch as we regard intellectually and emotionally rich lives as unconditionally good and intellectually and emotionally shallow lives as unconditionally bad. For a person with the normal range of intellectual, emotional, and physical capacities, it is a very bad thing to lead a simple and one-dimensional life with no opportunities for intellectual, emotional, and physical challenge or growth. One's life is made worse, not better, if, after informed and ideal deliberation, that is the sort of life to which one aspires.<sup>137</sup>

As applied to my theory (the no priority theory), Brink's objection would be this: My theory entails that A's welfare is partly a function of A's desires, but that is implausible, for A's desires are contingent occurrences that may or may not happen to track various goods (knowledge, friendship, health, etc.) that seem rather stably – or, as Brink might say, *unconditionally* – to be part of A's welfare.

I call this objection *the missing desires objection*. How should I respond to this objection? One move is to say (1) that this objection is correct in that it says that on the no priority theory the basic goods are not stable components of A's welfare (since on the no priority they are part of A's welfare only if, and only when, A happens to desire them), and then to say (2) that this isn't a problem (since it seems right to say that the basic goods aren't stable components of A's welfare). However, this seems like an inadequate response. After all, at least for the vast majority of us, it really does seem as though friendship, accomplishment, health, etc., are stable components of our own welfare. It would be very strange, for instance, to think that friendship is sometimes part of my welfare, and sometimes not part of my welfare. On the contrary, the truth seems to be that friendship is always part of my welfare.

But, given that the basic goods are stable components of almost everyone's welfare, doesn't it follow that the no priority theory is defeated by the missing desires objection? No, this doesn't follow. Indeed, the no priority theory can handle the missing

desires objection. I say this because human desire with respect to the basic goods (that is, with respect to friendship, health, etc.) is remarkably widespread and stable. Yes, there may well be some cases where individual humans lack the desire for one or more of these goods. But such cases are significantly fewer and farther between than is commonly thought; and, further, once we examine the details of these cases, we will see that these cases are not anything like decisive counterexamples to the no priority theory.

One point of clarification is in order here. The no priority theory as such says nothing about the remarkably widespread presence of, and the remarkable stability of, humans' desires for the basic goods. To argue that human desire with respect to the basic goods is remarkably widespread and stable (as I will do in sections 24-28) is to defend the no priority theory in a particular way. This particular defense of the no priority theory is, I believe, necessary: If the no priority theory isn't defended in this way, then it seems that the missing desires objection defeats the no priority theory.

The last thing to do, before jumping into sections 24-28, is to make a note of the following two distinctions. (1) Earlier (in section 8) I brought up a distinction between two sorts of cases: the actual/close to actual sort and the far from actual sort. By 'actual/close to actual cases' I intend non-fanciful cases. As for 'far from actual cases', I intend fanciful cases: Yes, we can imagine these cases, but we don't have any good evidence that they can occur in our world (or, more specifically, our environment) or in any world/environment that is relevantly like ours (say, a world/environment where the same laws of nature are operative, and where the history is substantially similar). Also, I noted (in section 8) that, when we are arguing about theories of human well-being, it seems helpful simply to bracket all far from actual cases. In sections 24-28 it will help if

we keep in mind that I am bracketing all far from actual cases. (2) Within cases of the actual/close to actual sort, we can distinguish between those that are standard and those that are non-standard. In speaking of standard cases, I mean to be ruling out, for example, cases involving people with psychological disorders, as well as cases involving people with severe brain damage. Also, if an otherwise normally functioning human is, say, extremely introverted and happens never or almost never to pursue friendship, then this case is a non-standard one. In sections 25-28 we will deal with non-standard cases. But in section 24 we will bracket all non-standard cases: We will assume that we are dealing *only* with standard cases. The goal for section 24 is to establish that, at least in the standard cases, the desires for the basic goods are *stably present* within humans.

#### **24. Standard cases**

Consider the basic goods – say, accomplishment, aesthetic experience, health, friendship, knowledge, and pleasure. Doesn't it sound strange – indeed, doesn't it sound like *a distortion* – to say of someone (for example, some close friend of yours) that she *just happens* to desire accomplishment, aesthetic experience, health, etc.? If someone desires accomplishment for herself, we certainly do not tend to say: 'Yet another *willy-nilly* desire of hers – no telling what will come next.' On the contrary, we tend to say: 'Not surprising – *who doesn't* desire accomplishment for herself?'. My claim here, then, is that it seems that the desires for the basic goods *just are* present within individual humans, and stably so – or, in other words, in such a way that they do not come and go, but rather are always there. Indeed, it seems that the desires for the basic goods are *built into* individual humans. After all, it seems that what would demand explanation is not

someone's desiring each of these goods for herself, but rather someone's failing to desire one or more of these goods for herself. *That* would be strange indeed. (Thus I am here appealing to a certain asymmetry. We expect, of any given human, that she has the desire for friendship, or for health, or for whichever basic good happens to be in question. Having this desire calls for no further explanation, but lacking it does.)

All of this can be further filled out. We ascribe desires to people as a way of explaining their actions, and it is obvious that people in general spend much of their lives pursuing instances of friendship, knowledge, health, accomplishment, aesthetic experience, etc. In explaining why this is so, it is sensible to assume that people in general have the desires for friendship, knowledge, etc., built into themselves.

One might object: 'To desire to write a good philosophy book is to desire an *instance* of accomplishment – but that isn't the same as desiring accomplishment. This matters. For, while it is true that people in general spend much of their lives pursuing instances of the basic goods, and while it is certainly right to infer that people in general desire instances of the basic goods, it isn't right to infer that people in general desire the basic goods.' Here, though, we should ask *why* people in general spend much of their lives pursuing instances of the basic goods. Consider an example: Someone gets on the subway so as to get to the museum; he goes to the museum because he wants to see some beautiful paintings; and he wants to see some beautiful paintings because he wants to experience beauty. Or again, this same person flies on a plane so as to get to the mountains; he goes to the mountains because he wants to see their beauty; and he wants to see their beauty because he wants to experience beauty.<sup>138</sup> These are perfectly natural chains of explanation for action, and each one terminates with a reference to a desire for



aesthetic experience. This person's desire for aesthetic experience helps to explain his various pursuits of, and desires for, instances of aesthetic experience. These various pursuits and desires are not sharply disconnected from one another. They have a certain unity/coherence to them, and this is so because they *flow out of* something more general, namely, this person's desire for aesthetic experience. And, speaking generally, this seems to be how we humans operate: Our various pursuits of, and desires for, instances of the basic goods seem to be best explained as *flowing out of* our desires for the basic goods.

Here one may protest: 'Your view is that A stably desires accomplishment, and that A's desires for various instances of accomplishment are best explained as flowing out of A's general desire for accomplishment. But do we really need to posit a general desire for accomplishment here? Isn't this extravagant? Shouldn't we just say that A desires various states of affairs that, for him, fall under accomplishment? Why do we need to add the controversial claim that A has a general desire for accomplishment? Wouldn't Occam's razor get rid of this general desire?'

This protest is unpersuasive. First, there really is nothing controversial about the claim that A has a general desire for accomplishment. To say that A has a general desire for accomplishment is to say that A has a disposition to bring it about that he engages in accomplishment. This claim is perfectly in keeping with common sense. In fact, the denial of this claim (or even agnosticism with respect to this claim) would be far more out of keeping with common sense than the acceptance of this claim. After all, in every normal case the agent in question just does have a general desire for accomplishment. Honestly, I don't know one person who doesn't desire accomplishment (or who isn't

disposed to bring it about that she engages in accomplishment). Second, Occam's razor doesn't get rid of A's desire for accomplishment. It is fine to say that, when it comes to desire-ascription, we should be parsimonious such that we should ascribe desires to an agent only if we absolutely must. But still, we have to explain everything that needs to be explained. The question, then, is that of what really demands explanation here. Just about everyone I know spends a great deal of their lives going about trying to get instances of accomplishment, friendship, pleasure, health, etc., for themselves. We must explain this fact. Here we might say that we desire instances of the basic goods *one-by-one, piecemeal, without any overarching unity involved*, etc. But this explanation makes the widespread and thoroughgoing human pursuit of instances of the basic goods seem to be more of a *happenstance* – more of a *mere coincidence* – than it really is. The better explanation, I believe, is the one that says that we desire various instances of the basic goods because our desires for various instances of these goods flow out of more general desires that we have – that is, more general desires that are simply *built into* us.<sup>139</sup>

Here is another objection to my view: 'You say that A has a general desire for accomplishment, and you say that A's desires for those things that are, for her, instances of accomplishment are desires that flow out of her general desire for accomplishment. Yet you also admit that there are cases where, although X instantiates accomplishment for A, A nonetheless has no desire for X. For instance, the case of the young academic fits this description. The young academic lacks the desire to finish her book, even though finishing her book instantiates accomplishment for her. How can you admit that A desires accomplishment in general for herself, that is, when there are things that instantiate accomplishment for A and yet that A is not at all disposed to bring about?'

I don't see a serious objection here. Someone can desire accomplishment in general for herself; her desires for various instances of accomplishment can flow out of this general desire; and yet it can also be true that, for *some* instances of accomplishment, it so happens that she has no desire for them. This might be because she doesn't know about these instances of accomplishment. Or, as in the case of the young academic, there may be a defective belief problem present. The young academic desires accomplishment for herself; she desires various instances of accomplishment for herself; and yet there is at least one instance of accomplishment – namely, finishing her book – that she does not desire. She doesn't desire this instance of accomplishment because her defective beliefs concerning her finishing her book have adversely impacted her desire-set: After reading the stinging criticisms from her colleagues and friends, she lost the belief that her book is any good (and this in turn caused her to lose her desire to finish the book). Nevertheless, the fact remains that finishing her book satisfies a desire that she has, namely, the desire for accomplishment. There is no reason to think that, just because she no longer desires a certain instance of accomplishment, it follows that she must not desire accomplishment in general. Speaking generally, moreover, I just don't see any significant worry for the no priority theory here: There seems to be no significant problem with the claim that the desires for *some* instances of basic goods can be missing even if, and even as, the general desires for the basic goods are all stably present.

Again, in order to understand my main point, there are two key things to see. One is that people in general spend much of their lives going about trying to get instances of accomplishment, friendship, aesthetic experience, health, etc., for themselves. The other is that, in explaining why this is so, it seems best to say that people's desires for these

instances of accomplishment, friendship, etc., flow out of more general desires that these people have – namely, the desires for accomplishment, friendship, etc.

All of this is meant to establish that, at least in the standard cases, humans have the desires for the basic goods built into themselves. If you still have doubts about this claim, then I urge you to consider yourself: Don't you desire friendship, accomplishment, aesthetic experience, etc.? And isn't it true that your desires for these goods are always there (or stably present)? Or, put differently, aren't you stably disposed to bring it about that you engage in friendship, that you engage in accomplishment, that you engage in aesthetic experience, etc.? Also, I urge you to consider others (say, people that you know well): Don't they desire friendship, accomplishment, aesthetic experience, etc.? And isn't it true that these desires are always there (or stably present)? Or, put differently, aren't these people stably disposed to bring it about that they engage in friendship, that they engage in accomplishment, that they engage in aesthetic experience, etc.? I strongly suspect that, if you are being honest, you will admit that you and others (say, those you know well) do indeed stably desire friendship, accomplishment, aesthetic experience, etc.

Earlier (in section 5) I noted that a desire is a mental state that one has: a mental state whereby one is disposed to bring about whatever it is that is desired. There I also noted that we need not feel or be conscious of our desires. With respect to our desires for the basic goods, I hold that, although these desires are stably present within us, we neither feel them nor are conscious of them, except perhaps occasionally. At least typically, then, we don't come to know about these desires through introspection of any sort. Rather, at least typically, we come to know about them by observing that people in general (including ourselves) spend much of their lives pursuing instances of knowledge,

health, etc., and then by judging that the best explanation for this is that people in general (including ourselves) have the desires for knowledge, health, etc., built into them.

One might wonder about the metaphysics that underlies all of this. But I will leave that open. Maybe humans in general have the desires for the basic goods built into them at least partly because of the way that evolution has gone.<sup>140</sup> And there are some other possible metaphysical explanations.<sup>141</sup> Speaking generally, this is a very important question, but it is also one that is best left for another inquiry. What matters for us here is just whether or not it is true that humans have the desires for the basic goods built into them. I have been arguing that, at least in the standard cases, this *is* true. As for what metaphysics best explains this, I will, to repeat, leave that open.

Before leaving this section behind, there is one more way in which I want to try to defend my claim that, at least in the standard cases, the desires for the basic goods just are stably present within individual humans. I am separating this argument from my previous remarks in this section because this argument seems to be distinct.

Ask yourself why is it that you spend so much of your life going for instances of accomplishment, friendship, etc. Isn't there something *more* involved here than your simply desiring various things that, for you, fall under accomplishment, friendship, etc.? Suppose that you actually fulfill a great many of your desires for various instances of, say, friendship. This would obviously satisfy you to a large extent. In other words, you would, to a large extent, have reached an equilibrium point – or a point of rest – with respect to your desire for friendship. But, to a significant extent, you *wouldn't* be at rest, desire-wise, with respect to friendship. What explains this? One explanation could be that there are still a great many instances of friendship that you desire, and yet that you do

not have – that is to say, you still have a great many unfulfilled desires for instances of friendship. However, it is doubtful that this is the *full* explanation here. For, aside from all of the instances of friendship that you desire and yet do not have, there seems to be something else that you desire and yet do not have – something else, that is, that goes above and beyond the various instances of friendship that you desire. The desire that I am speaking of here is the desire for friendship in general for yourself: the desire to have the whole thing, as it were. Getting tons of instances of friendship – even great instances of friendship such as being married to someone who you truly love, and who truly loves you – will not give you the whole of what you desire here. There is something more that you desire here: This more is your desire for friendship in general for yourself.

Similar remarks can be made with respect to the other basic goods. Getting various instances of these goods won't *fully* satisfy your desires with respect to these goods, and part of the explanation for this is that you desire more than just instances of these goods – you desire these goods themselves. (In speaking of one's not being fully satisfied on account of one's having instances of the basic goods rather than these goods themselves, I don't mean to be speaking of a lack of satisfaction, or an emptiness, that one necessarily feels or that one is necessarily conscious of. As I said above: Although the desires for the basic goods are stably present within us, I doubt that we either feel them or are conscious of them, except perhaps occasionally.)

The argument that I have been making here echoes the famous claim from Augustine that our hearts are restless until they rest in God. Here the suggestion from Augustine seems to be that our faculty of desire can't reach a point of rest, or a point of equilibrium, until we are united with an infinite good (namely, God). This is, of course,

an explicitly Christian claim from Augustine. But, for philosophical purposes, we can strip away the religious elements of this claim and still have on hand an important, non-religious insight into the nature of human desire – namely, the insight that there seems to be something *infinite* about human desire. (Note that the roots of the Christian view that human desire is for the infinite were in place long before Christianity came along – here, for instance, think of Plato’s *Symposium*.) For our purposes here, the really important point is that our desires with respect to the basic goods don’t seem to *reduce* simply to our desiring finite instances of these goods. It seems as though the having of finite instances of these goods can fulfill us to a significant extent – but it cannot *totally* fulfill us. The desire for more, the disposition to bring more about, will still be there. We are pushed, then, to the view that we must desire more than just tons and tons of finite instances of the basic goods – indeed, we must desire the basic goods themselves.

In spite of all that has been said, I suspect that some readers will still find the desires for the basic goods to be intolerably vague. The challenge that might be put to me here is this: ‘Can you provide us with a clearer characterization of what a desire for a basic good is?’. Honestly, though, I don’t think there is anything more that I can do here than to repeat what I have already said: For someone to desire, say, friendship is for her to have the disposition to bring it about that she engages in or experiences friendship. I see nothing intolerably vague in this. Also, it may matter who, in particular, is leveling the criticism against me here. OL theories are more convincing than DF theories, and so I would worry more about any criticism that might come at me from the OL theory camp. But, with respect to this particular criticism – the criticism that the desires for the basic goods are intolerably vague or nebulous – I don’t see why this is the sort of charge that an

OL theorist would level against me. OL theories require – indeed, are centered around – the basic goods. Therefore the criticism from OL theorists can't be that the basic goods themselves are intolerably mysterious entities – rather, the criticism has to be that *the desires for* the basic goods are what are intolerably mysterious. But why are these desires intolerably mysterious? Is it because these desires are aimed at very general things? But we desire very general things *all the time*.<sup>142</sup> For instance, there is nothing odd about our saying that so-and-so desires love, or joy, or whatever. Here our claim is that so-and-so is disposed to bring it about that she engages in or experiences love, or joy, or whatever. There isn't anything suspect about the desires that we are speaking of here. And the same holds for the desires for the basic goods – they are desires for very general things, and there is nothing suspect about them.

Stepping back now, if it is granted that, at least in the standard cases, humans have the desires for the basic goods built into them, then it follows that, at least with respect to the standard cases, the no priority theory can handle objections that assert that the no priority theory can't account for the fact that friendship, health, etc., seem to be stable components of one's welfare. What remains to be seen, though, is whether the no priority theory can still handle objections of this sort once we turn to non-standard cases. We will discuss non-standard cases in sections 25-28.

## **25. Non-standard cases not involving psychological disorders**

Before we start considering any non-standard cases, a quick reminder is in order: We are bracketing cases that are far from actual (or that are such that we have no good evidence that they can occur in our world/environment or any world/environment that is



relevantly like ours). Also, before considering any cases, I should note the following. We might say: 'If there is an actual/close to actual non-standard case that involves someone who seemingly never pursues friendship, then we should hold that this person must not desire friendship – or, at any rate, the burden of proof should be on those who claim that this person desires friendship, since, by all appearances, this person doesn't desire friendship.' But I disagree. If we consider people in general, it certainly seems as though they desire friendship, accomplishment, etc., and stably so. I *start*, then, by assuming that people have the desires for the basic goods built into them; and, for any given case, I will give this assumption the benefit of the doubt. For many actual/close to actual non-standard cases, then, my strategy will be to show that we can convincingly explain what is happening without giving up the assumption that the human in question (whoever she may be) stably desires all of the basic goods.

I do admit, though, that there are some actual/close to actual non-standard cases that involve humans who lack the desire for one or more of the basic goods. With respect to these cases, then, I will employ a different strategy: I will argue that we don't have compelling reasons for thinking that the non-desired basic good (whatever it may be) is part of the welfare of the human in question (whoever she may be).

We can begin by discussing the case of the extreme introvert. The objection from OL theorists here is something like this. 'Consider an extreme introvert who spends her days alone. Since she never or almost never pursues instances of friendship, she must not desire friendship. But, if she doesn't desire friendship, then, according to the no priority theory, friendship isn't part of her welfare. Yet that is implausible, for engaging in friendship surely makes her better off.' Is this case actual/close to actual or far from

actual? The answer depends on how the details of the case are filled in. Details should be filled in here, for there should be some sort of an explanation for why the extreme introvert doesn't pursue instances of friendship. If the details that are supplied are fanciful, then the case is far from actual and so can be bracketed. But, if the details that are supplied are non-fanciful, then the case is actual/close to actual and so must be considered. I contend that, if the details that are supplied are non-fanciful, then we will be able convincingly to explain the extreme introvert's non-pursuit of friendship without giving up the assumption that she desires friendship. But what would some non-fanciful details look like?

Here the following remarks are in order. Friends sometimes hurt (humiliate, betray, etc.) each other. In short, when we pursue friendship, we open ourselves up to being burned. Further, most of us, at least by adulthood, form more or less set routines (working, eating, sleeping, etc., routines). And, of course, we sometimes feel pressure from our spouses, children, relatives, and other friends to alter our routines. Maybe in the past our introvert has been badly burned by other people; and/or maybe she has extremely rigid routines and thus finds routine deviations very difficult. Also, it may be that our introvert places a high emphasis on, say, accomplishment. Indeed, it may be that she spends much of her life working (in solitude, apparently). And, naturally, if one is pursuing something besides friendship, then (unless this other thing happens to line up with friendship) one isn't going to be able to pursue friendship. The point here, then, is this. Even though our extreme introvert spends her days alone, it would be incautious of us to infer that she lacks the desire for friendship. After all, it may well be that, although she really does desire friendship, she doesn't pursue it because she is averse (1) to some

of its concomitants (say, to being vulnerable to being hurt by others and/or to changing her routines) and/or (2) to forsaking some other thing(s) for its sake (for instance, to forsaking accomplishment for its sake).

Some questions arise here. Why should we think that being vulnerable to being hurt by others is a concomitant of friendship? And why should we think that changing one's routines is a concomitant of friendship? Isn't it plausible to think that, rather than being concomitants of friendship, these things truly fall under friendship?

As far as ordinary language goes, yes, it is plausible to think that these things truly fall under friendship, and so aren't concomitants of it. However, by 'friendship' OL theorists intend something that is significantly *narrower* – or *more fine-grained* – than what ordinary language intends. (As I indicated in section 17, OL theorists use 'friendship' in a normatively/evaluatively loaded way.) Here consider an example. Say that I am attending a college football game with my friend Joe, that it is raining and cold, and that I am run down (physically). In this case the state of affairs *my spending time with my friend Joe* is an instance of friendship for me; and so, according to OL theorists, this state of affairs is an aspect of my welfare. However, the experience taken as a whole includes more than just the state of affairs *my spending time with my friend Joe*. For instance, the experience taken as a whole also includes the state of affairs *my spending time in cold rain when I am run down*. It may well be, then, that the experience taken as a whole isn't intrinsically good for me (or doesn't add to my welfare). One may quibble with some of the details here, but the general point is clear: Certain states of affairs do, and certain states of affairs don't, truly fall under friendship; and it is important for us to make the proper separations. Alfonso Gomez-Lobo is an OL theorist who discusses this

matter by bringing up the example of bacteria in a glass of water.<sup>143</sup> The bacteria aren't *internal* to the H<sub>2</sub>O molecules; they are *external* to them. Similarly, although instances of friendship are often surrounded by states of affairs that aren't intrinsically good for one, such states of affairs are *external* to instances of friendship – that is, such states of affairs don't truly fall under friendship.

Let's now return to our extreme introvert. Even if she is averse to being vulnerable to being hurt by others and/or to changing her routines, it doesn't follow that she is averse to friendship or to any of those *very narrow* (or *very fine-grained*) states of affairs that, for her, fall under friendship. All that follows is that she is averse to some concomitant(s) of friendship (or, in other words, to some states of affairs that *surround* those states of affairs that are, for her, instances of friendship). Moreover, with this last point in place, it makes sense to conclude the following. Our extreme introvert *does* desire friendship; and her non-pursuit of it results not from her being averse to it, but rather from her being averse (1) to some concomitants of it (say, to being vulnerable to being hurt by others and/or to changing her routines) and/or (2) to forsaking some other thing(s) for its sake (for instance, to forsaking accomplishment for its sake).

Actual/close to actual non-standard cases that involve people who seem never to pursue accomplishment can be handled similarly: Once the non-fanciful details of these cases are filled in, we will be able convincingly to explain these people's non-pursuit of accomplishment without giving up the assumption that they desire accomplishment. Here consider that many athletes must train painfully hard, many students must study painfully hard, many people in business must work painfully hard, etc., in order to get accomplishment for themselves. This raises the question: Is it worth it *on the whole* to

pursue accomplishment? Many people (in many cases) answer no, but that certainly doesn't imply that they lack the desire for accomplishment. After all, it is sensible to think that they do desire accomplishment, but that they don't pursue it because they are averse (1) to some of its concomitants (say, to fighting fatigue and stress) and/or (2) to forsaking other things for its sake (say, to forsaking time spent with friends for its sake).

But now consider that someone might say: 'I can agree with what you have said so far, but notice that you have been avoiding the hardest actual/close to actual non-standard cases – say, the ones involving psychological disorders. It is true that the no priority theory looks like the best theory of human well-being – that is, *if* we bracket all cases involving psychological disorders. However, once these cases are considered – and surely they must be considered – it becomes very hard to accept the no priority theory.'

In the next few sections (sections 26-28) we will discuss various cases involving psychological disorders. Naturally, my argument will be that the no priority theory can adequately handle these cases.

## **26. Non-standard cases: depression, eating disorders, and masochism**

Here in section 26 we will consider three psychological disorders: depression, eating disorders, and masochism. Our question with respect to depression is whether severely depressed individuals desire accomplishment; our question with respect to eating disorders is whether those with eating disorders desire health; and our question with respect to masochism is whether masochists desire pleasure.

*Depression:* Let's first try to get at least a rough handle on what it is like to be depressed. There are two (overlapping) categories of things that we should consider here:

(1) the abnormal levels of fatigue, irritability (and/or anger), stress, depletion, anxiety, dread, etc., that depressed people experience and (2) the distorted beliefs that depressed people have.

We can start with the first of these two categories. And here, in particular, we can start with two quotes. The first quote is from Paul Gilbert, a clinical psychologist:

Having worked with depressed people for many years and having heard many different stories, the thing that stands out is the sheer misery of it. Another way I encountered depression was personally, nearly twenty years ago...My depression was of the milder kind but even so it was associated with panic attacks, many sleepless nights, a terrible sense of having failed in important areas of my life, and a deep dread. Researchers rarely talk about dread, but I think it is a good word to sum up what people feel when they are depressed. I can't say I felt particularly sad, as you might if you lose a loved one, more frightened, irritable and joyless. Hence I would choose the word dread or 'living in dread' to describe it.<sup>144</sup>

The second quote is from *The Encyclopedia of Phobias, Fears, and Anxieties*. It states:

Two of the biochemicals that tend to be out of balance in depressed people are serotonin and nonrepinephrine. Scientists think that an imbalance of serotonin may cause the sleep problems, irritability, and anxiety many depressed people suffer. Likewise, an improper amount of nonrepinephrine, which regulates alertness and arousal, may contribute to the fatigue and depressed mood of the illness...Other body chemicals also may be out of balance in depressed people. Among them is cortisol, a hormone that the body produces in response to extreme cold, anger, or fear [cortisol is sometimes described as a stress hormone].<sup>145</sup>

Medications (for instance, anti-depressants) help many depressed patients by balancing out – or, in many cases, by replenishing – certain of their brain chemicals.<sup>146</sup>

The word 'replenishing' is apt, for the depressed person feels depleted (he feels drained of his normal zip or energy).<sup>147</sup> In severe cases this feeling of depletion can be so thoroughgoing that even performing the most minor of activities – getting out of bed, getting dressed, driving to work, and so on – can be terribly burdensome.<sup>148</sup> In short, depression is, physically speaking, a horrible state to be in – the depressed person simply feels terrible.

But, of course, depression is not just a matter of feeling terrible – or of feeling abnormally low. It is also a matter of having abnormal cognitive problems (or distorted beliefs). One might doubt that this is so. Michael Stocker states:

Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such "depressions" is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength.<sup>149</sup>

Also, with this passage from Stocker in mind, Michael Smith says: "It is a commonplace, a fact of ordinary moral experience, that practical irrationalities of various kinds – various sorts of 'depression' as Stocker calls them – can leave someone's evaluative outlook intact while removing their motivations altogether."<sup>150</sup> I honestly don't know exactly what Stocker and Smith (respectively) think about depression; and (at any rate) my aim is not to criticize either Stocker or Smith. Rather, my aim is just to make it clear that, contrary to what is sometimes thought, depression is *very much* a cognitive problem.

Here, then, we should discuss cognitive theories of, and cognitive therapies for, depression. And here, in particular, we should start with a quote from Steven Hollon:

With respect to depression, errors in thinking usually take the form of undue pessimism and lack of confidence in the self...Depressed individuals are troubled by negative automatic thoughts in specific situations (e.g., "I won't get that job even if I apply" or "She won't go out with me even if I ask her") that spring from more abstract underlying beliefs (e.g., "I'm incompetent" or "I'm unlovable") and are the source of such rules for living (underlying assumptions; e.g., "If I don't try I won't fail" or "If I don't get close to anyone I can't be rejected"). These beliefs are part of a larger cognitive schema that also includes the operation of logical errors...like all-or-none thinking or selective abstraction that serve to keep the depressed individual from recognizing the inaccuracy of his or her beliefs...Cognitive therapy for depression is based on the notion that correcting these erroneous beliefs and maladaptive information-processing strategies can reduce distress and facilitate adaptive coping...The approach has traditionally

focused on encouraging clients...to gather information and run experiments to test the accuracy of their beliefs...<sup>151</sup>

Hollon's above remarks about "more abstract underlying beliefs" are important. Why do I say that?

I do because cognitive theories of depression entail that the depressed person has certain *core dysfunctional beliefs*. These are very abstract beliefs to the effect that one is, say, helpless or unlovable. These beliefs are latent during life periods that are not very stressful. However, as Ari Solomon and David Haaga note:

[A.T. Beck's well known cognitive theory of depression proposes that these core dysfunctional beliefs] are easily reactivated by negative experiences that resemble the conditions under which the original beliefs were formed. For example, Joanne, who was orphaned at age seven, may have nearly forgotten her childhood belief that she is "all alone in the world" until her husband's request for a divorce brings it suddenly and painfully back into awareness.<sup>152</sup>

Further, once these core dysfunctional beliefs are reactivated, they act as *a negative filter* through which other cognitions pass. What happens, as Solomon and Haaga note, is this:

[A]ppraisals of specific day-to-day experiences (i.e., perception, interpretation, and recollection of experiences) begin to be systematically distorted to conform with negative belief. In other words, individuals begin to engage in *cognitive distortions* which render their experiences consistent with their underlying negative belief(s)...[A] depressed man who fails to please his wife sexually on one occasion may conclude that he is therefore destined always to displease her sexually. This illustrates the distortion known as *overgeneralization*....<sup>153</sup>

Solomon and Haaga then proceed to note other cognitive distortions that depressed people engage in – for example, *magnification* ("overestimating and overattending to the negative aspects of one's experiences") and *personalization* ("relating negative events to oneself without evidence").<sup>154</sup>

Two further points are in order here. (1) There is much empirical evidence that suggests that cognitive theories of depression are true. As Solomon and Haaga state:



There is extensive research support for the description of depressive thinking embedded in Beck's cognitive theory [here it should be said that Beck's is the most famous and most established cognitive theory of depression]. Depressed people, regardless of clinical subtype, exceed nondepressed people, including those with other psychological disorders, in negative thinking about themselves, their personal worlds, and their prospects for the future.<sup>155</sup>

(2) There is also much empirical evidence that suggests that cognitive therapies for depression are effective. As Hollon states:

It appears that cognitive therapy is about as effective as medication (even for more severely depressed patients) and that it has an enduring effect that reduces subsequent risk. This enduring effect is perhaps its most important feature; there is strong and consistent evidence that teaching people how to examine the accuracy of their own beliefs can reduce their subsequent risk for depression, whether they have a history of prior depression or are just approaching the age of risk. Given that depression tends to be a chronic recurrent disorder and that medications only suppress the expression of symptoms (they do little to change the underlying cause of the disorder), this is a very major advance.<sup>156</sup>

We have now said enough, I think, by way of describing depression. It is time, then, for us to return to our main question: Should we hold that the severely depressed person has the desire for accomplishment? Or, instead, should we say that she lacks this desire?

Here consider the following. Those of us who don't have depression sometimes feel pretty bad. We rarely, if ever, experience that distinctive sense of *dread* (and/or *being worried sick*) that so pervades the depressive's psyche. But, that said, we certainly do sometimes struggle with fatigue, stress, irritability, anxiety, etc. Indeed, facing these sorts of difficulties is part of normal living. Just think of days where you hardly slept the night before or of days where you are sick (say, with a very bad cold). If you can function at all on these sorts of days, you can do so only in a perfunctory (or relatively lifeless) way. On days such as these, you surely *do* desire instances of accomplishment for yourself. It is not as though your desire for accomplishment goes away just because you are having one of those days where you have to drag yourself around. However, on

days such as these, you are not as likely (as you usually are) to act so as to fulfill your desire for accomplishment. After all, there are (on days such as these) other desires that are much more pressing. Examples of these much more pressing desires might be the desire to recuperate (physically) and the desire simply to avoid further pain (say, further pain that might be incurred by performing even minor activities like getting out of bed, getting dressed, etc.). Now imagine that, instead of having to drag yourself around for just a short stretch, you have to do this for an extended period of time (say, for months on end). Very roughly, that is what it is like, physically speaking, to be depressed.

But, of course, it should be noted that there are some crucial cognitive differences between a non-depressed person who drags herself around and a depressed person who does the same. (1) The non-depressed person believes that her feeling low is a temporary and conquerable phenomenon. But the depressed person tends to see his feeling low as going on indefinitely and insuperably. Indeed, the depressive doesn't see any end in sight to his condition. In an effort to make himself feel normal again, he has already tried doing what worked in the past, and (like many depressives) he may have even tried to carry out some superstitious plans – but all to no avail.<sup>157</sup> (Here the depressive might be overlooking the fact that, provided that he can afford it, he might get help – that is, medication and therapy. For some reason, many depressives simply don't think of getting help.<sup>158</sup>) (2) The non-depressed person believes in (or has faith in) the general efficacy of her own actions. Even if she is feeling horrible and can do no more than drag herself around, she has confidence that doing so will at least to some extent help her in bringing about what she desires. For instance, the non-depressed college student believes that going to her afternoon class today (say, even though she has a very bad cold) will

pay off. Indeed, she thinks that, although it will be extremely unpleasant, it will nonetheless help her in keeping up with the course material. Not so with the depressive, though – he tends to view his actions as being largely ineffective in bringing about what he desires. For instance, the depressed college student tends to view his going to his afternoon class today not only as something that is extremely unpleasant, but also as something that is largely an exercise in futility (say, because he believes that he is generally incompetent and that he is probably going to fail the course in the end anyway).

To get to my point: Depression doesn't extirpate one's desire for accomplishment. For instance, the depressed college student still desires to do well academically and, more generally, still desires accomplishment. 'But, if the depressive desires accomplishment, then why doesn't he pursue it?' It makes sense to think that the depressed person's non-pursuit of accomplishment results from (1) his acting on desires that are significantly more pressing than his desire for accomplishment (if he is like many depressives, then he will often be acting on a rather pressing desire to avoid further pain – say, further pain that might be incurred by performing even minor activities like getting out of bed) and (2) his believing that, even if he were to pursue instances of accomplishment, his actions would likely be ineffective in helping him to get such instances.

*Eating disorders:* There are different kinds of eating disorders, but here it will suffice, I think, if we focus on anorexia and bulimia. There are serious health problems associated with anorexia and bulimia: inability to menstruate, muscle cramping, fatigue, dizziness, loss of energy, constipation, inability to concentrate, hair-loss, inability to stay warm, etc.<sup>159</sup> Anorexics and bulimics don't *desire* to have any of these health problems. Indeed, they are *averse* to them, and this is evidenced by their *worrying* about them

(and/or by their being *frightened* by them). Here consider a quote from a therapist who treated an anorexic named Tracy:

Tracy came to therapy at the age of 19, in her sophomore year of college... Tracy stood 5'6" tall and weighed about 98 pounds. She had started losing weight deliberately about two years earlier, although she had been thin even then, weighing about 120 pounds... When she came to me Tracy was frankly unhappy about the prospect of gaining weight, but *she was also frightened by* feelings of fatigue, the loss of her period for more than six months, and her difficulty in being able to concentrate on schoolwork [the italics are mine].<sup>160</sup>

Here we might add that anorexics and bulimics are often motivated to get better in large part precisely because they are averse to the health problems that they have. Take Tracy again: "*Her fatigue and poor concentration*, which she correctly understood as the effects of malnutrition, *helped motivate her* to stop restricting food [the italics are mine]."<sup>161</sup>

Given that anorexics and bulimics don't want – indeed, are averse to – the health problems that are associated with their condition, why do they allow themselves to fall into (and/or to remain in) their condition? Roughly, the answer is this. Anorexics and bulimics want health, but they also strongly want to be very thin.<sup>162</sup> Further, they have *distorted beliefs* about eating, weight gain, and their own bodies; and, along with their desire to be very thin, these distorted beliefs lead them to fall into (and/or to remain in) their condition.

With regard to these distorted beliefs, we might consider Tracy again: "Tracy took more than one and a half years to reach her target weight [of 122 pounds]. She would gain and lose. For a long time, she described her food intake as 'huge,' when in fact she was eating under 2,000 calories per day (on which she could not gain)."<sup>163</sup>

Or, again with respect to these distorted beliefs, we might consider the following case about a bulimic named Jane:

[T]he clinician emphasized the importance of eating regular meals and snacks and worked with Jane to develop a plan to improve her nutritional intake...[Jane] was skeptical that eating regular meals and snacks would lead to weight gain. The clinician emphasized that most normal-weight individuals do not gain weight as a result of eating regular meals and snacks...[T]he clinician explained to Jane that the vomiting associated with bulimia nervosa is ineffective in ridding the body of the food ingested during a binge. Although Jane was convinced that she ‘got all the food out’ while purging, the clinician explained that the body absorbs a portion of the food eaten during a binge eating episode...The clinician suggested a technique called hypothesis testing, in which she and Jane would ‘test’ her belief. They considered possible ways of experimenting to see whether Jane’s belief that she would definitely gain weight eating regular meals was true. They agreed to a one-week trial period, in which Jane would eat according to the meal plan...When Jane weighed herself one week later, she noted that her weight had not increased. In addition, she was struck by how much less she was binge eating and vomiting. Jane discovered that eating regular meals and snacks and increasing the amount of food she was consuming made her much less hungry and less likely to binge in the afternoon.<sup>164</sup>

The case of Jane illustrates that cognitive therapy can help those with eating disorders – with the guiding idea here being that, once the agent’s distorted beliefs about eating, weight gain, and her own body are corrected, she will be able to recover from her condition.

Just how successful is cognitive therapy in treating those with eating disorders? Carol Peterson, Ph. D., and James Mitchell, M.D., answer this question as follows: “The research literature suggests that CBT [cognitive-behavioral therapy] is the first treatment of choice for bulimia nervosa...CBT’s role in the treatment of anorexia nervosa is less clear, although it appears promising for the treatment of this condition as well.”<sup>165</sup> (For some reason, only a few studies have been done on CBT’s effectiveness in treating anorexia – these studies indicate that CBT is indeed effective in treating anorexia, but, of course, more studies should be done in order to confirm these results.<sup>166</sup>)

Let’s now sum up our answer to the question ‘Do anorexics and bulimics desire health?’. Anorexics and bulimics do want to be healthy. Their health problems (inability

to menstruate, loss of energy, etc.) result not from their desiring to have them, but rather from (1) their having and acting on the desire to be very thin and (2) their having distorted beliefs about eating, weight gain, and their own bodies.

At this point, however, it is very natural to object (roughly) as follows. ‘The anorexic, say, desires to be *very* thin – indeed, *abnormally* thin. So she obviously lacks the desire to be of a healthy weight. But being of a healthy weight is itself a large part of being healthy. So how can you be so sure that the anorexic desires health?’

But this objection is confused. Weight-wise, the anorexic wants to be very thin. However, she still wants to be in the range of healthy and normal. She doesn’t want to be in the range of disordered, unhealthy, abnormal, emaciated, etc.<sup>167</sup> ‘But why, then, does she allow herself to become (and/or to remain) abnormally thin, emaciated, etc.?’ The problem here is *cognitive*. When *we* look at the anorexic, we correctly see (or believe) that she is abnormally thin, emaciated, etc. But that isn’t what *she* sees (or believes). From her own point of view, she is (weight-wise) either right/close to right (that is, very thin, but still in the range of healthy and normal) or maybe even overweight. Sadly, one of the main symptoms of anorexia is (to quote a group of experts on anorexia) “disturbance in body image, for example, seeing one’s self as fat even when one is bone thin or emaciated.”<sup>168</sup>

*Masochism*: Do masochists desire pleasure? Yes, they do. In this sense they are normal. Masochists are abnormal, however, in that they pursue, and perhaps even desire, pain. Sexual masochists, for example, want to experience the pleasure that sex brings. In this sense they are normal. They are abnormal, though, in that they willingly engage in sexual practices where they are made to experience pain (for instance, where they are

whipped). Why do they do this? They apparently derive pleasure from doing this – indeed, they apparently derive pleasure from pain (that is, in specific situations).<sup>169</sup>

It is sometimes said that (impossible though it is) masochists experience pain as pleasure. With respect to this matter, the psychologist Roy Baumeister states:

One couple I interviewed suggested that the pain becomes pleasant, but when I asked them how this happens they quickly added that the sensation of pain remains painful and unpleasant. The pain is tolerable and there is something about it that appeals to the masochist, but it is simply not true that pain turns into pleasure.<sup>170</sup>

Here Baumeister is thinking of pleasures and pains as feelings (or sensations); and he is also thinking – no doubt rightly – that a feeling of pain can't turn into (or be experienced as) a feeling of pleasure. It might be possible (for all I know) for one to have a *feeling-judgment complex* where (1) the feeling part of this complex is positive (a pleasure) and (2) the judgment part of this complex is directed toward a negative feeling (a pain) that is *separate from* (or *not part of*) this complex. And, when masochists say that pain turns into pleasure (or that they experience pain as pleasure), it may be that they are talking about a (positive) feeling-judgment complex that is directed toward a pain – say, a pain that they feel in virtue of being whipped. Here, then, there wouldn't be *one* feeling that is experienced first as a pain and then as a pleasure – rather, there would be *two* feelings, one being a pain and the other being (roughly put) a pleasure taken *in* a pain. Feelings do not have intentionality, but we speak as if they do when we say that we take pleasure *in* (or that we enjoy/like) such-and-such. When we say things like this, we may be referring to feeling-judgment complexes that are directed toward something.

At any rate, there is no need to say any more here. The fundamental point is just this. Since masochists *do* desire pleasure, they pose no threat to the no priority theory.

## 27. Non-standard cases: autism

With respect to the actual/close to actual non-standard cases discussed in sections 25 and 26, my argument was that we can convincingly explain what is happening without giving up the assumption that the human in question (be she the extreme introvert, or the depressive, or the anorexic, or whatever) is really just like everybody else in that she stably desires all of the basic goods. However, not all actual/close to actual non-standard cases can be handled in this way. Here, for instance, we might consider autism.

*Autism:* The main question here is whether autistics desire friendship. We know that autistics have serious social know-how problems. High functioning autistics such as Temple Grandin and Donna Williams have written about this matter. Grandin states:

I am like the lady referred to as S. M. in a recent paper by Antonio Damasio...She has a damaged amygdala. This part of the brain is immature in autism. S. M. has difficulty judging the intentions of others, and she makes poor social judgments. She is unable to recognize subtle changes in facial expression, which is common in people with autism...I have learned how to understand simple emotional relationships that occur with clients. These relationships are usually straightforward; however, emotional nuances are still incomprehensible to me...I still have difficulty understanding and having a relationship with people whose primary motivation in life is governed by complex emotions, as mine are guided by intellect. This has caused friction between me and my family members when I failed to read subtle emotional cues...My job and career make life worth living. Intellectually complexity has replaced emotional complexity in my life.<sup>171</sup>

In reading Grandin and Williams it is clear that they sometimes avoid other people because they believe that, if they were to interact with others, this would result in awkwardness/friction.<sup>172</sup> As Williams notes, “fear” and “defensiveness” often lead her to reject social invitations to hang out with others.<sup>173</sup> Further, in a letter to her doctor, Williams writes of the difficulties that she has in making friendships; and she ends the letter by saying: “Can you help me with this because I am slowly accepting that I *want* to make *real* friendships, and it hurts that I can’t get past step one?”<sup>174</sup> The general point



here is that, with respect to those autistics who seem to show little interest in friendship, there is good reason to think that at least many of them do desire friendship and would pursue it and engage in it in a substantial way if not for their grave social know-how problems.

One recent scholarly article on autism notes: “Although many [autistic] individuals report that they have challenges when it comes to emotions, it is not an absence of interest in emotional interaction. They continue to try to relate to others.”<sup>175</sup> Here, then, the suggestion seems to be that autistics do desire friendship, even if the appearances may suggest otherwise. And this is definitely true with respect to autistics such as Grandin and Williams: They obviously do desire friendship and do engage in it, at least to a very limited extent.<sup>176</sup> However, as compared to many autistics, autistics such as Grandin and Williams are extremely proficient in terms of language and social capacities. And, for the no priority theory, the worry is that it is plausible to think that there are some severe autistics who are born without and never acquire any appreciable desire for friendship, not even after substantial treatment efforts are made. Let’s refer to these autistics as *friendship-less autistics*. It is hard to know whether there are any actual friendship-less autistics, but, whether there are or not, it seems that I must admit that the case of friendship-less autistics is actual/close to actual. I say this because, based on our observations of the severe social deficits of certain actual autistics, it seems likely that there at least can be friendship-less autistics who live in our world/environment and in worlds/environments that are relevantly like ours.

In dealing with friendship-less autistics, my key claim is this. While it is true that they have no appreciable desire for friendship, it is *also* true that they have no appreciable

*receptivity* for friendship. In order to understand what I am thinking here, it will help if we consider a quote from Laura Schreibman, one where she discusses the social deficits of severely autistic children:

It is not uncommon to hear a mother report that as an infant her [autistic] child did not hold up his arms in anticipation of being picked up, did not look at her when held, or was “stiff” or “rigid” to hold...The parent may describe the slightly older child as not wanting to be held, cuddled, or kissed, sometimes actively resisting or avoiding expressions of affection or other social overtures. Typically the child is not upset when the parent leaves or particularly happy when the parent returns after an absence; he seems, in fact, not to notice. Children with autism usually do not come to the parent for comfort if frightened or injured, nor are they likely to be consoled by the parent’s efforts to comfort them. The parents are not used as a “secure base” when the children are in new or strange settings. Rather, the parents have to be very careful to keep the child close because if they become separated, the child most likely will not be distressed nor try to locate the parent...Many parents come to feel that their child does not love them or need them as people, but simply relates to them as objects that provide what the child wants...As one might expect from someone who is not socially involved, individuals with autism typically fail to show empathy or to understand the feelings of others...Autistic children are as unresponsive to their peers as to their parents. Other children are typically ignored or actively avoided.<sup>177</sup>

Do the severely autistic children that are discussed in this quote benefit from the love/friendship that their parents offer them? It isn’t clear that they do, for, sadly, it seems that, at least most of the time, these children are *impervious* to the love/friendship that their parents offer them.

This is significant, for it suggests that, with respect to friendship-less autistics, we don’t have compelling reasons for thinking that friendship is part of their welfare. Think of the point this way. The anorexic has the receptivity for health. If she eats healthfully, she will gain weight and so be better off. Thus health is part of her welfare. And the same sort of thing holds for the extreme introvert: If she engages in friendship, she will benefit from it. But matters are different for the friendship-less autistic: If he hangs out with other people, then this will do him no appreciable good, for he can’t appreciably

absorb/take in friendship, and so can't appreciably benefit from it. Thus, even though he lacks the desire for friendship, it doesn't follow that he lacks the desire for something that is part of his welfare.

In saying that friendship-less autistics can't appreciably absorb/take in friendship, I am thinking that, in order truly to engage in and benefit from friendship, you must have the ability to link the thoughts and/or emotions of others to yourself in such a way that you are appreciably emotionally moved by these others. Friendship-less autistics lack this ability and thus lack the ability truly to engage in and benefit from friendship.

In invoking the notion of movability here, I am drawing on an article by Peter Hobson ("On Being Moved in Thought and Feeling: An Approach to Autism"). Hobson says:

The central idea is that *being moved by others* – and here I am talking about movement in subjective orientation, especially as these involve feelings and attitudes – is one of the most significant features of human social life. It is foundational for experiencing people *as* people with their own subjective orientations to the world....<sup>178</sup>

Hobson discusses various studies that show that autistics are severely impaired with respect to the ability to be moved by others. With regard to one of these studies, Hobson states:

The research involved us interviewing children and adolescents with autism, along with closely matched participants without autism, to discover how they think about themselves... The principal group difference was that very few of the statements made by the adolescents with autism referred to social relations, and many of those that did merely implied social awareness (for example, with reference to themselves as 'good' or 'nice'). Half as many of the adolescents with autism mentioned their families, but most striking of all, not a single individual with autism mentioned a friend except in passing (and even this happened only once), whereas 70 per cent of the non-autistic adolescents did so... Now these findings suggest that even within their own minds, individuals with autism rarely think about themselves in relation to other people who have attitudes toward them, and rarely take up a stance in which they think about others, think about

themselves, and compare the two in relation to one another. Such forms of intrapsychic role-taking may well implicate being moved to the position of others, and then moving back to a focus on oneself.<sup>179</sup>

All of the studies that Hobson discusses illustrate that severe impairment with respect to being moved by others (for instance, in thought and in emotion) is something that is basic to autism. With regard to friendship-less autistics, the important point to take away here is that their impairment with respect to being emotionally moved by others is, as it were, total. This, then, is why friendship-less autistics can't truly engage in and benefit from friendship.

I have said that friendship-less autistics cannot benefit from friendship because they have no appreciable receptivity for it. However, I should stress there is a notable possible sense in which friendship is part of the welfare of friendship-less autistics. For a friendship-less autistic, friendship isn't part of his welfare, at least not in the same way that it is for a normally functioning human. Yet there is a closer relationship between friendship and the welfare of a friendship-less autistic (on the one hand) than there is between friendship and the welfare of, say, a carrot (on the other hand). If a friendship-less autistic's constitution were to change such that he remained a human being but no longer had severe autism (and so did desire friendship and was able appreciably to absorb/take in friendship), then friendship would be part of his welfare. In that possible sense friendship is part of his welfare. But the same doesn't hold for a carrot, for we can't imagine a carrot's constitution changing such that it remains a carrot *and* has the desire for, and the capacity to absorb/take in, friendship. (Put another way: Although friendship isn't a near-possible component of the friendship-less autistic's welfare, friendship is a far-possible component of the friendship-less autistic's welfare. In

contrast, for a carrot, friendship is not only not a near-possible component of its welfare – friendship is also not a far-possible component of its welfare.)

To sum up: Even though friendship isn't part of the welfare of friendship-less autistics, there is a key possible sense in which friendship is part of the welfare of these autistics. This point matters, since it makes it clear that the no priority theory need not say what seems false, namely, that friendship holds no prudential importance at all for friendship-less autistics. If you look at any given human being and ask yourself whether there is a relationship of fit between friendship and his well-being, then surely the answer cannot be that there is no relationship of fit whatsoever. That answer, after all, is simply too flat. Nevertheless, in the case of friendship-less autistics, it would be wrong to hold that the relationship of fit between friendship and their welfare is the same in kind as the relationship of fit between friendship and the welfare of some (any) normally functioning human being. Thus arises the need for finding a principled middle ground here. The no priority theory can say that, although friendship isn't a near-possible component of the welfare of friendship-less autistics, friendship nonetheless is a far-possible component of the welfare of friendship-less autistics; and, fortunately, this seems to get the matter right.

One final note here: We might try to imagine cases that involve autistics who lack the desire for friendship, and yet who can appreciably benefit from friendship; and it might be thought that these imagined cases (if they really can be imagined) are a problem for the no priority theory. But in fact these imagined cases do not threaten the no priority theory, for they are not actual/close to actual – rather, they are far from actual. I say this because our observational evidence of actual autistics gives us no reason to think that there can be autistics of this sort (that is, of the sort that lack the desire for friendship, and

yet that can appreciably benefit from friendship) in our world/environment and in worlds/environments relevantly like ours.

## **28. Non-standard cases: psychopathy**

We will now focus our attention on psychopaths, who are often thought of as being callous and inhuman. Our main question with respect to psychopaths is the question of whether they desire friendship.

*Psychopaths:* We should start by trying to get a handle on what psychopathy is (or on what the psychopath is). Robert Hare (a psychologist) states:

Most clinical descriptions of the psychopath make some sort of reference to his egocentricity, lack of empathy, and inability to form warm, emotional relationships with others – characteristics that lead him to treat others as objects instead of as persons and prevent him from experiencing guilt and remorse for having done so. After an extensive review of the literature, McCord and McCord...concluded that the two essential features of psychopathy are *lovelessness* and *guiltlessness*.<sup>180</sup>

One other characteristic of note here: The psychopath is impulsive.<sup>181</sup> His impulsive behavior often gets him into trouble (as evidenced by the fact that our prisons are full of psychopaths).

With respect to friendship, the psychopath often *appears* to engage in it. Here consider Hare's description of a psychopath named Donald:

By all accounts Donald was considered a willful and difficult child. When his desire for candy or toys was frustrated he would begin with a show of affection, and if this failed he would throw a temper tantrum; the latter was seldom necessary because his angelic appearance and artful ways usually got him what he wanted...When he was 22 he married a 41-year-old woman whom he had met in a bar. Several other marriages followed, all bigamous. In each case the pattern was the same: he would marry someone on impulse, let her support him for several months, and then leave. One marriage was particularly interesting. After being charged with fraud Donald was sent to a psychiatric institution for a period of observation. While there he came to the attention of a female member of the

professional staff. His charm, physical attractiveness, and convincing promises to reform led her to intervene on his behalf. He was given a suspended sentence and they were married a week later. At first things went reasonably well, but when she refused to pay some of his gambling debts he forged her name to a check and left. He was soon caught and given an 18-month prison term.<sup>182</sup>

Thus Donald often *appears* to fall in love and in general to care about others. Indeed, he often *appears* to engage in friendship. But this is just an appearance – it is not the real thing. Indeed, the psychopath’s engagement in friendship is superficial. Notice, moreover, that it is not just that the psychopath *doesn’t* engage in true friendship – it is also that he *can’t* engage in true friendship. As Hare (here following Cleckley) states:

According to Cleckley, the psychopath lacks the ability to experience the emotional components of personal and interpersonal behavior – he mimics the human personality but is unable to really *feel*. Thus, although his verbalizations (for example, “I’m sorry I got you in trouble”) appear normal, they are devoid of emotional meaning....<sup>183</sup>

The psychopath, then, neither does nor *can genuinely care* about others and their welfare. By extension, then, he is incapable of desiring, and also of engaging in, true friendship.

The above remarks provide us with a description of the psychopath. Or, to be more precise, they provide us with a description of what I will refer to as the *pure* psychopath. One question to consider here is this: Are there any actual *pure* psychopaths – or, instead, is the notion of the *pure* psychopath merely a clinical construct? Here we might consider the following. The Hare Psychopathy Checklist-Revised (PCL-R) is the standard tool that psychologists (and other mental health experts) use in order to diagnose psychopathy.<sup>184</sup> Hare and some of his colleagues state:

The PCL-R uses a semistructured interview, case history information, and specific scoring criteria to give an individual zero, one, or two points on each of twenty items. Thus, total scores can range from zero to forty. The higher the score, the more closely an individual matches the prototypical psychopath...[A] threshold score of thirty has proven useful for a research diagnosis of psychopathy.<sup>185</sup>

To quote Hare and company, the first eight of the twenty items in question here are as follows: “(1) glibness/superficial charm; (2) grandiose sense of self-worth; (3) need for stimulation/proneness to boredom; (4) pathological lying; (5) conning/manipulative; (6) a lack of remorse or guilt; (7) a shallow affect; (8) callous/lack of empathy...”.<sup>186</sup> And the next twelve items are all in the same ballpark. The main point to take away here is this. Reflection upon how mental health experts diagnose psychopathy suggests that there are no actual *pure* psychopaths – there are only some actual people who we refer to as psychopaths because they are relatively close to the idea that we have of what a *pure* psychopath would be (that is, were there actually to be one).

Here, moreover, we should consider two recent studies. In their 2006 article (“Psychopathic, Not Psychopath: Taxometric Evidence for the Dimensional Structure of Psychopathy”) John Edens and some of his colleagues ask whether psychopathy is “dimensional or categorical in nature.”<sup>187</sup> That is to say, they ask: “[T]o what extent does the construct of psychopathy identify a fundamentally distinct class of individuals who differ qualitatively from the rest of society?”<sup>188</sup> And, in light of their research on 876 prison inmates and court-ordered substance abuse patients (all of whom were administered the PCL-R), they conclude that psychopathy is dimensional, not categorical:

Of course, our findings do not imply that distinctions between psychopaths and nonpsychopaths cannot be made for practical purposes, such as risk management or violence prediction. If other research teams replicate our findings, however, this would imply that such distinctions are purely pragmatic and do not “carve nature at its joints,” because in the case of psychopathy, there are no discrete joints to carve.<sup>189</sup>

This study’s findings have been confirmed. In their 2007 article (entitled “A Taxometric Analysis of the Latent Structure of Psychopathy: Evidence for Dimensionality”) Hare and some of his colleagues state that, in light of their research on 5,408 male prison inmates



in various North American institutions, it appears clear that psychopathy is dimensional, not categorical. Indeed, Hare and company assert that “the observed results [of their research] support a dimensional structure for psychopathy and its components...”.<sup>190</sup>

These two studies are obviously not infallible indicators that psychopathy is dimensional. But they do strongly suggest that it is. And, if psychopathy is dimensional, then it stands to reason that there are no actual *pure* psychopaths – rather, there are only some actual people who we refer to as psychopaths because they are relatively close to the idea that we have of what a *pure* psychopath would be (that is, were there actually to be one).

Another point worth noting here: If we knew that psychopathy is completely a function of some specific gene(s) and/or some specific brain part(s), and if we knew that some individual humans are completely deficient (or completely abnormal) in respect of the gene(s) and/or brain part(s) in question, then it might make perfect sense to think that there are some actual *pure* psychopaths. But in fact we don’t know exactly causes psychopathy. It is thought that one’s genetic make-up matters here and, in addition, that one’s being abused as a child may matter here.<sup>191</sup> Further, Hare and company state:

Recent psychobiological and neuro-imaging research indicates that altered brain functioning may underlie the behavioral and cognitive anomalies associated with psychopathy. In particular, psychopaths appear to have difficulty in modulating goal-directed behavior because of a failure or defect in automatic mechanisms for attending to and evaluating peripheral cues. They also appear to have difficulty in understanding and processing emotional information; emotion does not play the same motivating and directive role for them as it does for normal individuals.<sup>192</sup>

At any rate, the point here is just this. Reflection upon our current knowledge of what causes psychopathy doesn’t push us toward the conclusion that there are some actual *pure* psychopaths.

One might think that, even if there are no actual pure psychopaths, the case of *pure* psychopaths is nonetheless actual/close to actual. The idea here would be this: As long as it is clear that there are some actual humans who are very close to being *pure* psychopaths, this gives us (or might give us) sufficient evidence for the claim that there at least can be *pure* psychopaths in our world/environment and in worlds/environments that are relevantly like ours. I doubt that the case of *pure* psychopaths really is actual/close to actual. But, even if I am wrong about this, it doesn't follow that the no priority theory is in trouble. For the no priority theory can handle *pure* psychopaths in the same general way that it handles friendship-less autistics. The idea here is this. *Pure* psychopaths lack the emotional capacities (and, in particular, the sympathy and empathy capacities) that a human being must have if he is to be able to engage in true friendship. Indeed, *pure* psychopaths are incapable of engaging in true friendship. With this point in place, it is sensible to conclude that *pure* psychopaths can't appreciably benefit from friendship – which is to say that friendship isn't a near-possible component of the welfare of *pure* psychopaths. However, as with friendship-less autistics, we should note that friendship is a far-possible component of the welfare of *pure* psychopaths. This point matters, for it would be wrong to claim that friendship holds no prudential importance at all for *pure* psychopaths. At any rate, the bottom line here is that *pure* psychopaths are not a problem for the no priority theory. For, even though it is true that *pure* psychopaths do not desire friendship, it doesn't follow that they lack the desire for something that is part of their welfare – after all, as we have just said, friendship isn't part of their welfare.

I have said that I doubt that the case of *pure* psychopaths really is actual/close to actual. One major reason that I doubt this is that, when we closely examine psychopaths,

we actually find that they are far less callous – and far more human – than the surface appearances suggest. Here we should think about the sickest people that we know of – say, serial killers like Jeffrey Dahmer. Certainly these people are very close to being pure psychopaths, aren't they? These people don't desire friendship, do they? They are just too callous, too inhuman, to desire friendship, right? Here we should bring up what Willem H.J. Martens refers to as the hidden suffering of the psychopath.<sup>193</sup> Martens says:

Like anyone else, psychopaths have a deep wish to be loved and cared for. This desire remains frequently unfulfilled, however, as it is obviously not easy for another person to get close to someone with such repellent personality characteristics. Psychopaths are at least periodically aware of the effects of their behavior on others and can be genuinely saddened by their inability to control it...Despite their outward arrogance, inside psychopaths feel inferior to others and know they are stigmatized by their own behavior...They see the love and friendship others share and feel dejected knowing they will never take part in it...Psychopaths are known for needing excessive stimulation, but most foolhardy adventures only end in disillusionment due to conflicts with others and unrealistic expectations. Furthermore, many psychopaths are disheartened by their inability to control their sensation-seeking...Although they may attempt to change, low fear response and associated inability to learn from experiences lead to repeated negative, frustrating, and depressing confrontations, including trouble with the justice system...As psychopaths age they are not able to continue their energy-consuming lifestyle and become burned-out and depressed, while they look back on their restless life full of interpersonal discontentment.<sup>194</sup>

Thus it seems that, although actual psychopaths (that is to say, actual people who would clinically qualify as psychopaths) are abnormal in certain respects, they are quite normal in that, just like anybody else, they *do* desire friendship.

But still, what about the really extreme – or really sick – clinically diagnosed psychopaths? Surely they don't desire friendship, do they? Here we should again turn to Martens:

Social isolation, loneliness and associated emotional pain in psychopaths may precede violent criminal acts...As psychopathic serial killers Jeffrey Dahmer and Dennis Nilsen expressed, violent psychopaths ultimately reach a point of no return, where they feel they have cut through the last thin connection with the

normal world. Subsequently their sadness and suffering increase, and their crimes become more and more bizarre...Dahmer and Nilsen have stated that they killed simply for company...Both men had no friends and their only social contacts were occasional encounters in homosexual bars. Nilsen watched television and talked for hours with the dead bodies of his victims; Dahmer consumed parts of his victims' bodies in order to become one with them: he believed that in this way his victims lived further in his body...For the rest of us it is unimaginable that these men were so lonely – yet they describe their loneliness and social failures as unbearably painful. They each created their own sadistic universe to avenge their experiences of rejection, abuse, humiliation, neglect and emotional suffering.<sup>195</sup>

Thus it seems clear that even the sickest of clinically diagnosed psychopaths *do* desire and *do* pursue friendship – though, admittedly, they pursue it in very strange ways (Martens notes that Nilsen even “wrote poems and spoke tender words to the dead bodies [of his victims], using them as long as possible for company”).<sup>196</sup>

In another article Martens and George Palermo (writing together) note the following about Dahmer:

At age 18, while driving his parents' car, Jeffrey gave a ride to a young man who was walking along the road stripped to the waist. Jeffrey was impressed by his looks, and he invited him to his home for a drink with the hope of having sex with him. When the boy refused his advances and wanted to leave, Jeffrey strangled him and sexually abused him. He later stated that he could not stand the idea of being abandoned and he experienced an irresistible desire to keep the boy with him. He later went on to kill 16 more victims, all adolescents and young men. As his obsession grew, he began saving body parts. He wanted to remember the appearance of his victims, and he took pictures of the corpses. They belonged to him. He exercised total control over his victims to the point that he attempted to make zombies out of some of them: He gave them alcohol in which sleep-inducing drugs were dissolved, and when they were half-asleep, he drilled holes in their skulls and injected muriatic acid into them to liquefy the brain matter, but when his experiment failed, he murdered them. The skulls of the six victims on whom he had perpetrated the horrendous act were found in his apartment at the time of his apprehension...As in the case of Dennis Nilsen, Britain's notorious and gruesome serial killer, Jeffrey Dahmer wanted company. Whether his urge to kill was driven by a longing for simple companionship or sex, Dahmer killed for it. He stated that he had wanted to be with his victims, to keep them with him, and he described his fear of being abandoned by them. He shared with many other serial killers not only a deep, violent, destructive hostility but also boredom, loneliness, fear of rejection, and an ambivalent craving for human closeness.<sup>197</sup>

To say the least, there is something seriously wrong with (and abnormal about) the way in which Dahmer pursued friendship. But that he desired it seems rather clear.

To sum up: The actual people who clinically qualify as psychopaths – even the very sickest ones – do seem to desire friendship. As such is the case, it is hard to see why reflection upon psychopathy should lead us to doubt the no priority theory.

Some final comments on actual/close to actual non-standard cases are now in order. I have discussed the case of the extreme introvert, the case of the depressive, the case of the anorexic, etc. No doubt there are cases that I did not discuss, but that might be discussed. My general point, though, should be quite clear by now. We have good reason to think that the no priority theory isn't threatened by actual/close to actual non-standard cases. In particular, we have good reason to think that one of the two general strategies that I have employed will work whenever we are dealing with an actual/close to actual non-standard case. The first strategy is to show that, contrary to what we might think, the human in question (be she the extreme introvert, or the depressive, or the anorexic, or whatever) really does desire the basic good in question (be it friendship, or accomplishment, or health, or whatever). If this strategy does not work, then the second strategy will. The second strategy consists in two moves. First, I point to a lack of receptivity for the non-desired basic good (whatever it may be), and in turn I argue that we don't have compelling reasons for thinking that this non-desired basic good is part of the welfare of the human in question. Second, I emphasize that, even though this non-desired basic good isn't part of this human's welfare, there remains a notable possible sense (namely, the far-possible sense) in which this basic good is part of this human's welfare.

## 29. The no priority theory and babies' desires

There are two more matters to discuss here in Chapter 3. In this section I will make some brief remarks about the desires of babies. Then (in section 30) I will make some comments about the relationship between prudential value and perfectionist value.

Sometimes OL theorists claim that, when it comes to babies, their view has a clear edge over DF theories. The guiding thought here is that the desires of babies aren't fit to play any serious role in determining the welfare of babies. Naturally, if this thought is on target, then the no priority theory is in trouble. Here we can consider something that Kraut says:

In fact, however, there is no plausibility in the idea that a baby's good can be constructed out of what she wants. Babies must, for their own good, be nurtured in a way that gives them certain competencies and brings them into certain human relationships; but the states of affairs that are good for them are not already present to them as the content of their desires. They do not know what is good for them and cannot know. So the conative approach to well-being has no resources for accounting for the good of infants. They do not make plans or decisions; and although they do have desires and preferences, these are not robust and complex enough to provide an adequate account of what is good for them.<sup>198</sup>

Kraut says that babies' desires are not robust and complex enough to provide an adequate account of what is good for them. But how complex, really, do babies' desires need to be in order to play a key role in determining their well-being?

The well-being of babies is not very complex – not in comparison to adults, anyway. Babies spend their days eating, sleeping, lying on little blankets, smiling, crying, shaking little toys that make noise, drooling, etc. These are not very complex lives that babies lead. To say this isn't in any way to downplay the moral or normative status of babies, but it is to state that the welfare of babies isn't very complex (relative to the welfare of adults). Since the well-being of babies is not very complex, babies' desires

don't need to be very complex if they are going to play a key role in determining the welfare of babies. In short, what is needed is a parallelism between the complexity of the content of A's desires and the complexity of the content of A's welfare. As long as this parallelism is there, the claim that A's desires are complex enough to play a key role in determining A's welfare is unproblematic.

Another point to consider here: In order for the no priority theory to work, it need not be true that babies desire every state of affairs that is part of their welfare. All that must be true is that babies desire the basic goods (friendship, health, etc.). And, on reflection, it does seem that we have good evidence in favor of the claim that babies desire the basic goods. If we observe babies, we can see (1) that they pursue health inasmuch as they eat, sleep, etc., (2) that they pursue friendship inasmuch as they cry for their parents to hold them, (3) that they react favorably to music and soothing voices, which suggests that they desire aesthetic experience, and (4) that there is within babies a disposition to accomplish things, as is evidenced by their straining to perform new tasks, such as clasping hands, rolling over, and (later) walking and talking.

One might be inclined to protest that the basic goods are so abstract, or so general in nature, that it is hard to believe that babies have the cognitive capacities required to desire the basic goods. But I don't see a serious objection here. It is common, within the philosophical literature on desire, to assume that babies can desire rather general things. For instance, while discussing the desires of babies, Schroeder states: "In common parlance babies want full stomachs, dry bottoms, cuddling, colorful sights, and soothing sounds, and these basic desires do not vanish as they grow older, though they gain a certain degree of sophistication."<sup>199</sup> And later Schroeder adds:

Human infants appear to be born with a large stock of intrinsic desires...They desire food and water, warmth, dry skin, human contact, and so on...From this basic stock of desires, new desires are acquired until as adults they have intrinsic desires for and aversions to a panoply of things.<sup>200</sup>

In these quotes Schroeder is assuming that there is no problem with the claim that babies are born desiring rather general things – things like full stomachs, soothing sounds, human contact, and so on.<sup>201</sup> These desires may not be very sophisticated, content-wise. But there is no reason to doubt that they are present. I admit that more could be said about the desires of babies. But I will not say any more here. In short, I don't think that the no priority theory needs to worry, in any serious way, about the desires of babies.

### **30. The relationship between prudential value and perfectionist value**

My view of the relationship between prudential value and perfectionist value may already be clear from what I have said in some of the above sections (see sections 2, 16, 19, and 22). Still, it may not be a bad idea if I now state/restate my view of this matter.

The no priority theory says that X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (health, friendship, etc.) and (2) X's obtaining satisfies some desire that A has (with the idea being that, if A does not desire X itself, then A must at least desire the basic good that X here instantiates for A). OL theorists conceive of the basic goods as prudential goods – that is, as prudential goods that are intrinsic, desire-independent, and relational. I agree with OL theorists that the basic goods are intrinsic, desire-independent, and relational. But on my view we shouldn't conceive of the basic goods as prudential goods. Yet, given that we shouldn't conceive of the basic goods as prudential goods, how should we conceive of them? There are two general ways to answer this question. (1) I might say that we



should conceive of the basic goods as non-prudential, intrinsic, desire-independent, and relational goods. And then I might say that this is the end of the story – that there is no need to put a specific label on the sort of value that is being invoked here. (2) I might put a specific label on the sort of value that is being invoked here. If I must do this, then I am willing to use the label ‘perfectionist value’.

But is it sensible to claim that perfectionist goods are non-prudential, intrinsic, relational, and desire-independent? Nobody will quibble with the claim that perfectionist goods are intrinsic goods (that is, non-instrumental goods) or with the claim that perfectionist goods are relational goods. Moreover, because perfectionist goods are for someone *qua* her being a human *as such* (and not *qua* her being the *individual* human she is), it stands to reason that an individual’s own desires do not play any direct role in making things be intrinsically good for her in the perfectionist sense. The only remaining question, then, is whether it is sensible to think of perfectionist goods as non-prudential goods. What, then, should we say here?

The concept of perfectionist value for humans is the concept of a sort of value that accrues to individual humans on account of their functioning properly/excellently as the sorts of things they are – that is, as human beings. In contrast, the concept of prudential value for humans is the concept of a sort of value that accrues to individual humans on account of their lives going well for them as the individuals they are. It seems, then, that we have two concepts here. Nevertheless, even if perfectionist value and prudential value are conceptually distinct, it doesn’t necessarily follow that they are metaphysically distinct. As I noted in section 16, there are plenty of OL theorists – for instance, Murphy, Finnis, and Kraut – who hold (or who at least seem to hold) that, even if perfectionist

value and prudential value are conceptually distinct, prudential value and perfectionist value are nonetheless metaphysically identical.

Why would anyone think that there is a metaphysical identity relationship here? Here is one way that someone might answer, or start to answer, this question. ‘Consider the states of affairs that we intuitively think of as being aspects of people’s welfare. These appear to be *the exact same* states of affairs that we intuitively think of as being perfective of people. Isn’t it true that engaging in friendship, that gaining knowledge, that accomplishing things, that having aesthetic experiences, and so on are not only aspects of people’s welfare, but are also perfective of people? It seems, then, that prudential value and perfectionist value never really come apart. This lends support to the claim that we are dealing with a metaphysical identity relationship here.’

Here we might consider two quotes from Kraut. Although Kraut doesn’t like the term ‘perfectionism’ (he prefers to speak of flourishing), he uses ‘flourishing’ to mean ‘functioning well’ (so he is talking about perfectionist value).<sup>202</sup> Here is the first quote:

A good theory of well-being should be built on a root idea that is obvious, widely recognized, and rich in implications. Clearly, flourishing is a good thing – good for what is flourishing. We can talk about a flourishing or thriving business or legal practice, but flourishing is primarily a biological phenomenon: “flower” and “flourish” are cognates. Above all, it is plants, animals, and human beings that flourish when conditions are favorable. They do so by developing properly and fully, that is, by growing, maturing, making full use of potentialities, capacities, and faculties that (under favorable conditions) they naturally have at an early stage of their existence. Anything that impedes that development or the exercise of those mature faculties – disease, the sapping of vigor and strength, injuries, the loss of organs – is bad for them.<sup>203</sup>

And here is the second quote:

To test the developmental theory of well-being (so we will call it – or “developmentalism” for short), we should of course consider the kind of living beings in whom we are most interested: human beings. What is good for them is determined by what it is for them to flourish – not what the means to their

flourishing are, but what constitutes their flourishing, that is, what its components are. To know what constitutes their flourishing, we must know something about human beings in general. We noted, in the previous section, that their flourishing requires the development not merely of physical powers, but of psychological powers as well. Which powers? Using the categories of common sense, we can say at least this much: a flourishing human being is one who possesses, develops, and enjoys the exercise of cognitive, affective, sensory, and social powers (no less than physical powers). Those, in broadest outline and roughly speaking, are the components of well-being.<sup>204</sup>

In light of these two quotes, we can see that Kraut holds that A's flourishing and A's faring well are one and the same (or are identical). What should we think of Kraut's view – that is, the view that perfectionist value and prudential value are identical? I think that, all things considered, it is best to reject this view. Two sets of points are in order here.

One: I am not completely convinced that perfectionist value and prudential value never come apart (or always inhere in the exact same states of affairs).<sup>205</sup> I have no problem with the claim that every state of affairs that adds to one's welfare is also such that it perfects one. To engage in accomplishment, knowledge, etc., doesn't just seem to add to my well-being – it also seems to perfect me. After all, it seems that, if I engage in accomplishment, knowledge, etc., then I thereby function well as a human. So far, then, there is no problem with the claim that perfectionist value and prudential value never come apart. Still, this claim may not be true. Here consider a quote from Steven Wall:

An objective account of the good need make no reference to the good of human beings. Works of art or aspects of the natural world might have value even if no human being existed. Perfectionist accounts of the good are accounts of the human good. They seek to identify the goods that contribute to the best or most perfect life for human beings...The best life for human beings can be understood in at least two importantly different ways. On the first understanding, such a life is understood in terms of well-being. The best life for a human being is a life that goes maximally well for the person who leads it. On the second understanding, the best life for a human being is understood in terms of excellence or success. An excellent human life need not be the one that is best in terms of well-being, for it is possible that the most excellent life that a human being can live requires him to make sacrifices in his own well-being for the sake of other persons or goods.

Thus the notion of an excellent human life is broader than the notion of well-being. For this reason, a general characterization of perfectionism should employ it rather than well-being.<sup>206</sup>

In this quote Walls says that “the notion of an excellent human life is broader than the notion of well-being.” Here Wall seems to be thinking that perfectionist value and prudential value can come apart (because perfectionist value is broader than prudential value). In particular here, the thought seems to be that there are cases where one makes a sacrifice in terms of one’s own welfare for the sake of advancing another’s welfare, and that in such cases one perfects oneself, but doesn’t thereby prudentially benefit.

I am wary of straightforwardly asserting that perfectionist value and prudential value can come apart. But I do think that there *might* be cases where one is perfected, and yet where one doesn’t prudentially benefit. To be clear about my position here: In my terminology, ‘the basic goods’ simply refers to the desire-independent, relational goods that OL theorists typically put on their lists (accomplishment, friendship, etc.). This leaves open the question of how we should conceive of the basic goods. Unlike OL theorists, I do not conceive of the basic goods as prudential goods. I am willing to conceive of the basic goods as perfectionist goods. This means that I am willing to say that every basic good is a perfectionist good. But that leaves open the question of whether every perfectionist good is a basic good. On reflection, I think that there *might* be some perfectionist goods (say, generosity and justice – and moral virtue in general) that do not belong on the list of basic goods. (Here I am referencing generosity and justice – and moral virtue in general – because Wall’s rationale for saying that perfectionist value is broader than prudential value is that we sometimes perfect ourselves inasmuch as we make sacrifices in our own well-being for the sake of advancing the

well-being of others. Such self-sacrificing action would seem to be generous or just action – or, in general, morally virtuous action.)

Do humans in general stably desire moral virtue (including generosity and justice) in the same way that humans in general stably desire friendship, health, etc.? I will leave this open. I accept (1) that accomplishment, friendship, health, pleasure, knowledge, and aesthetic experience count as basic goods, and (2) that humans in general stably desire these goods. Also, I am willing to conceive of these goods as perfectionist goods, and I am confident that, excepting for very rare cases, these goods are stable components of people's well-being (because they are basic goods that are stably desired). Beyond this, however, there are certain things I am leaving open – for instance, (1) the question of whether moral virtue (which is a perfectionist good) is a basic good and also (2) the question of whether humans in general stably desire moral virtue.

All along, I have been claiming that X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (health, knowledge, accomplishment, etc.) and (2) X's obtaining satisfies a desire that A has. But, if I were pressed, I would be willing to say: X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of a perfectionist good and (2) X's obtaining satisfies a desire that A has. One follow-up question here would be this: Which perfectionist goods do humans in general stably desire? I would answer: Humans in general stably desire accomplishment, friendship, knowledge, health, pleasure, and aesthetic experience. *Maybe* humans in general also stably desire some other perfectionist goods, since *maybe* humans in general also stably desire generosity and justice (or, more generally, moral virtue). Again, I will leave this matter open.

Two: Even if it is true that perfectionist value and prudential value never come apart (or always inhere in the exact same states of affairs), we should still deny that perfectionist value and prudential value are identical. If X is intrinsically good for me in the perfectionist sense, then X is intrinsically good for me *qua* my being the *kind* of thing I am, namely, a human being. But, if X is intrinsically good for me in the prudential sense, then X is intrinsically good for me *qua* my being the *individual* that I am. Thus it seems that prudential value is *personal* in a way that perfectionist value isn't. Because there is something personal about prudential value that distinguishes it from perfectionist value, it is sensible to think that the *for* part of 'good for' in the case of prudential value must be filled out differently than the *for* part of 'good for' in the case of perfectionist value. In particular, it is sensible to think that a reference to A's own mental states/pro-attitudes is necessary when it comes to the spelling out of 'good *for* A in the prudential sense', though not when it comes to the spelling out of 'good *for* A in the perfectionist sense'. As I noted in section 19, many philosophers hold that prudential value is personal in a way that perfectionist value isn't. Here we might consider something Rosati says:

I will generally use the expression 'personal good' in place of more common terms, such as 'flourishing', 'welfare', 'well-being', and 'self-interest'... The goodness at issue is nonmoral, nonaesthetic, and nonperfectionist, although an individual's good may include as components moral, aesthetic, and perfectionist values.<sup>207</sup>

Thus Rosati goes so far as to use 'personal good' instead of 'well-being', 'welfare', or 'prudential value'.

Also, even if perfectionist value and prudential value always inhere in the exact same states of affairs, it may be that they are not identical to each other and that this can be seen in that they differ in *the extents to which* they inhere in certain states of affairs. I

might affirm that this holds for my own welfare theory, and in partially filling this out I might say: Given that X is both perfective of A and an aspect of A's welfare, we should hold that the extent to which X is perfective of A doesn't depend on A's desires, while (in contrast) the extent to which X is an aspect of A's welfare depends on (or is a function of) both the extent to which X is perfective of A and the strength with which A desires X. And, if I were to say this, I could further fill things out in such a way as to make it clear that X might be perfective of A to a greater or lesser extent than X is prudentially good for A. (To be clear: I am not claiming that perfectionist value and prudential value inhere in certain states of affairs to differing extents – this is just something that I *might* claim.)

It is worth adding that, even if we bracket my view of well-being and focus solely on objectivism about well-being, there may be good reasons for thinking that, even if perfectionist value and prudential value always inhere in the exact same states of affairs, they nonetheless differ in the extents to which they inhere in certain state of affairs. In elaborating on this point, it will help, I think, if I very briefly discuss a certain debate that takes place among natural law OL theorists.

Finnis is a natural law OL theorist who maintains that perfectionist value and prudential value are identical and that there is no objective hierarchy among goods such as moral virtue, friendship, knowledge, aesthetic experience, etc.<sup>208</sup> Yes, an individual may rank, say, friendship over knowledge in his or her own life, but the fact remains (thinks Finnis) that, objectively, friendship and knowledge are equally important. Resistance to Finnis sometimes arises among natural law OL theorists (1) because, for both Aristotle and Aquinas, there is an objective hierarchy such that, say, knowledge and moral virtue are ranked above things like aesthetic experience, health, and friendship, and

(2) because some who work within the natural law theory tradition think that Aristotle and Aquinas are right to think that some goods just are objectively more important than other goods.<sup>209</sup>

What I want to point out here is this. If a natural law OL theorist is thinking in perfectionist terms, then it may well be right to think that there is an objective hierarchy among the goods in question. For instance, it may well be right to think that in order for someone (anyone) to function properly/well as a human being, she must give moral virtue some sort of priority over goods such as accomplishment and aesthetic experience. However, if a natural law OL theorist is thinking in prudential terms, then (in line with what Finnis would say) it probably isn't right to think there is any objective hierarchy among goods such as moral virtue, health, friendship, aesthetic experience, etc. As I noted in section 19, the perfectionist viewpoint is strictly third-personal, while the prudential viewpoint is essentially first-personal (though I am willing to say that the prudential viewpoint includes the third-personal, perfectionist viewpoint as one of its moments). And, if it is admitted that the perfectionist viewpoint is strictly third-personal and that the prudential viewpoint is essentially first-personal, then the idea of there being an objective ranking of goods for perfectionist value seems plausible, while the idea of there being an objective ranking of goods for prudential value seems less plausible. If these remarks are correct, then perhaps the natural law OL theorist should say this. The same goods – moral virtue, accomplishment, health, etc. – matter for both perfectionist value and prudential value. And, indeed, these two sorts of value always inhere in the exact same states of affairs. But the sort of weight that a given good gets in perfectionist terms may well differ from the sort of weight that this same good gets in prudential



terms. Here, then, a natural law OL theorist would be advancing a view according to which perfectionist value and prudential value are *not* identical to one another (and where this is so even in spite of their always inhering in the exact same states of affairs).

Summarizing my general point here: I deny that perfectionist value and prudential value are identical. But, if perfectionist value and prudential value always inhere in the exact same states of affairs, then this denial may seem implausible. What I have been trying to show, however, is that, even for certain OL theorists who believe that perfectionist value and prudential value always inhere in the exact same states of affairs, there may be good reasons for denying that perfectionist value and prudential value are identical.

Although I deny that perfectionist value and prudential value are identical to one another, it is worth emphasizing that my view of this matter is actually very close to the view that Murphy, Finnis, and Kraut take. I say this because on my view all of us (or, in any case, almost all of us) are made better off in prudential terms by the exact same states of affairs that make us better off in perfectionist terms – or, to be more precise, this holds at least if we are bracketing moral virtue. It is likely that, out of all of those philosophers who deny that perfectionist value and prudential value are identical to one another, I am the one among them whose position is the closest to the position that Murphy, Finnis, and Kraut take.<sup>210</sup>

To conclude Chapter 3: The main point of this chapter has been that, in contrast to DF theories and OL theories, the no priority theory adequately captures both the *good* and the *for* parts of ‘good for’ (that is, of ‘*prudentially* good for’). We should now move onto Chapter 4, where we will discuss hybrid theories in general.

## Chapter Four: Hybrid Theories in General

### 31. Brief remarks on the main hybrid theory options

All hybrid theories entail that human welfare is a function *both* of some sort of non-prudential value *and* of some sort of pro-attitude. Naturally, the idea here is (1) to employ some sort of a value condition in order to capture the *good* part of ‘good for’ and (2) to employ some sort of pro-attitude condition in order to capture the *for* part of ‘good for’. However, this leaves open the question of just what sort of value condition, and just what sort of pro-attitude condition, a hybrid theorist should rely on. With respect to the value condition, all hybrid theories rely on a sort of value that is intrinsic, non-prudential, and pro-attitude independent – the open question is that of whether to rely on a *relational* or a *non-relational* sort of value. As far as the pro-attitude condition goes, it seems that *desire*, *enjoyment*, and *cognitive endorsement* are the main pro-attitudes that a hybrid theorist might seize upon.

The no priority theory is, of course, a hybrid theory. With respect to the value condition, the no priority theory relies on a *relational* sort of value. And, with respect to the pro-attitude condition, the no priority theory relies on *desire*. Other hybrid theorists may object either to the no priority theory’s value condition or to the no priority theory’s pro-attitude condition (or to both). What we need to do, I think, is to consider various value plus pro-attitude combinations, all with the aim of seeing why the combination that the no priority theory relies on is the best possible combination. In the next section (section 32) we will consider the possibility of relying on the pro-attitude of *enjoyment*. Enjoyment might be combined with either a relational or a non-relational sort of value. I don’t think, though, that it matters which sort of value enjoyment is combined with,

since, either way, enjoyment does not seem to be a promising pro-attitude to rely on. This point will become clear shortly. In section 33 we will consider the possibility of relying on the pro-attitude of *cognitive endorsement* – which might be combined with either a relational or a non-relational sort of value. I am going to argue that relying on cognitive endorsement is a bad idea. In section 34 we will turn to the pro-attitude of *desire* – which also might be combined with either a relational or a non-relational sort of value. I will argue that the no priority theory combination – that is, the desire-cum-relational value combination – is indeed the best possible combination for a hybrid theorist to rely on. As for section 35, that will conclude our entire inquiry.

### **32. Should we rely on enjoyment?**

As I have said, the no priority theory relies on desire. But, in advancing a hybrid theory, one might think that enjoyment is a more promising pro-attitude to rely on.

Robert Adams has recently said things that suggest that he may well be a hybrid theorist who relies on the pro-attitude of enjoyment. In his book *Finite and Infinite Goods* Adams gives excellence – which is a non-relational, intrinsic sort of goodness – pride of place. The following quote from Adams brings this point out:

Within the realm of what is good for its own sake, and not just instrumentally good, most contemporary ethical thought focuses mainly on *well-being* or welfare – that is, on the nature of human flourishing or what is good *for* a person. The theory developed here, however, gives primary place to *excellence* – the type of goodness exemplified by the beauty of a sunset, a painting, or a mathematical proof, or by the greatness of a novel, the nobility of an unselfish deed, or the quality of an athletic or philosophical performance. It is the goodness of that which is worthy of love or admiration, honor or worship, rather than the good (for herself) that is possessed by one who is fortunate or happy (though happiness may also be excellent, and worthy of admiration).<sup>211</sup>

Later, after criticizing DF theories of well-being, Adams states:

If desire-satisfaction theories of the nature of a person's good won't do, what will? Without pretending to offer here a complete theory of the nature of a person's good, I wish to explore the idea that what is good for a person is a *life* characterized by *enjoyment of the excellent*. More precisely, I shall argue that the principal thing that can be noninstrumentally good for a person is a life that is hers, and that two criteria (perhaps not the only criteria) for a life being a good one for a person are that she should enjoy it, and that what she enjoys should be, in some objective sense, excellent.<sup>212</sup>

In light of what Adams says, we can see that there is a hybrid theory that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is excellent and (2) A enjoys X. Let's refer to this hybrid theory as *Adams's theory*.

I don't know whether Adams himself actually accepts what I am referring to as Adams's theory. There are multiple plausible interpretations of Adams's various comments concerning human welfare. On one plausible interpretation Adams is an OL theorist who holds that there is just one basic good (namely, the enjoyment of the excellent), while on another plausible interpretation Adams simply hasn't made up his mind about *exactly* what sort of view of human welfare he accepts. And there may be some other plausible interpretations available. At any rate, I find it convenient to use the label 'Adams's theory' for the above mentioned theory, and so I will go ahead and do so.

In order to assess whether Adams's theory is convincing, we need to get a handle on what enjoyment is. Enjoyment is best thought of as being a pro-attitude. But this actually makes things complicated. After all, when one hears the word 'enjoyment', one immediately tends to think of a positive *feeling* (or a positive *sensation*). However, feelings, taken on their own, lack intentionality. This means, then, that feelings, taken on their own, can't be pro-attitudes. After all, pro-attitudes necessarily take objects. But what, then, is enjoyment? If it can't just be a positive feeling, then what is it? It seems that in ordinary language the word 'enjoyment' refers – or at least paradigmatically refers

– to something that is felt, even if only in a very mild way. Thus it is misleading, I think, if we use ‘enjoyment’ in philosophy as a technical term that denotes something that need not be felt. With regard to the question of what enjoyment is, I don’t have a worked out answer. But here is a proposal that we might consider. To enjoy X is to have a *judgment-feeling composite* (or complex) that is directed toward X – with the idea here being that the judgment part of this composite has intentionality, while the feeling part of this composite is a positive feeling/sensation that lacks intentionality. If A enjoys watching some football game, then we might say that A has a judgment-feeling composite that is directed toward his watching this game – where the judgment part of this composite consists in A’s judging/believing that he is watching this game, and where the feeling part of this composite is a positive feeling of A’s. I don’t know how accurate this proposal is. But, regardless, the main point here is not that this proposal is correct – rather, the main point is that, whatever enjoyment is, it must involve a positive feeling.

It is possible, of course, that someone can have a pro-attitude toward X and yet not feel anything positive. For instance, I can *desire* X and yet not feel anything positive. Or again, I can *cognitively endorse* X and yet not feel anything positive; that is to say, I can judge/believe X to be valuable in some way or other (for instance, I can judge/believe X to be an aspect of my welfare) and yet not feel anything positive. But, again, it really doesn’t seem right to say that I can *enjoy* X and yet not feel anything positive.

Now let’s ask: Is Adams’s theory – the theory that says that well-being consists in (and only in) the enjoyment of the excellent – plausible? The answer is no. I say this because we often get things that are aspects of our welfare, and yet do not feel any positive feelings in the process. It follows, then, that one need not enjoy X in order for X

to be an aspect of one's welfare, which means that Adams's theory fails. Here, of course, the critical question is whether it really is true that we often get things that are aspects of our welfare, and yet do not feel any positive feelings in the process. The following case will illustrate that, yes, this really is true.

Suppose that some philosopher is working through some of the nitty-gritty details of a certain paper of hers, and suppose that she is not enjoying herself. She is laboring through the day, and, although what she is writing is actually quite good, she is not having a pleasant time, not at all. The work is just too hard – too dense and cumbersome – for her really and truly to *enjoy* what she is doing. Also suppose, however, that she is accomplishing quite a bit (thought-wise, or philosophy-wise) on this day of work. Shouldn't we say, then, that the accomplishment that she gets for herself on this day adds to her welfare, even though she is not really and truly *enjoying* what she is doing? Yes, she *desires* the accomplishment that she is getting for herself. Further, she *cognitively endorses* the accomplishment that she is getting for herself: She judges/believes that this accomplishment is a good thing, and she judges/believes that this accomplishment is part of her welfare. And yet it seems wrong to say that she really and truly *enjoys* this accomplishment. After all, in order to enjoy this accomplishment, there would need to be a positive feeling present. But there is no positive feeling present. Still, even though she doesn't enjoy this accomplishment, it seems that it would be wrong to deny that this accomplishment is part of her welfare. Naturally, this is just one case. But, that being said, it is not an isolated case. There are many others that illustrate the same general point. The bottom line here, then, is that it seems that we should reject any and all welfare theories that entail that X cannot be an aspect of A's welfare unless A enjoys X.

What might be said in response here? One move is to say the following. ‘Yes, it is true that we sometimes get things that are aspects of our welfare without feeling any positive feelings in the process. For instance, it is true that the philosopher in the above example gets something that is part of her welfare – in particular, some accomplishment – and yet doesn’t feel any positive feelings in the process. However, once we properly understand the nature of enjoyment, it will be clear that there is room for us to say that this philosopher really does enjoy the accomplishment that she is getting here.’

In order to understand this response better, we might consider some of Adams’s comments about the nature of enjoyment. Adams says:

[E]njoyment must be understood here as including more than what is usually meant by ‘pleasure’. Enjoying life is not simply a matter of “feeling good” or having pleasant experiences. It is also, and much more, a matter of the zest or interest with which one engages in the activities of life.<sup>213</sup>

Here Adams seems to say that, although enjoyment does essentially involve a positive feeling, it essentially involves more than just a positive feeling. However, Adams soon seems to suggest that enjoyment may not essentially involve a positive feeling. He says:

There is yet another point, however, at which an obsession with pleasure might mislead us about the kind of enjoyment or liking of one’s life that is essential to a person’s good; and that is the relation of enjoyment to time. If we think of enjoyment in terms of feelings of pleasure, it will be natural to think that what matters is that the moments of one’s life should be enjoyed or liked while they are occurring. That certainly adds to one’s enjoyment of life, if it happens, but it is not the only thing to be taken into account when considering to what extent someone has enjoyed her life. Suppose that she has succeeded in swimming the English Channel. Perhaps the hours she spent in the water were mostly unpleasant, full of weariness, anxiety, and cold. Nonetheless, we may count her swimming the channel as something that she enjoys in her life, if she savors the achievement.<sup>214</sup>

And Adams goes on to say:

Savoring the achievement is important. It is not so much while she was doing it as after she had done it, and knew that she had succeeded, that she liked the swim.

And the value that she sets on it retrospectively seems, intuitively, quite relevant to how good it was for her that her life included this episode...More important than the duration and intensity of the moments of retrospective pleasure, when we are considering whether she enjoyed her life more by making the swim, is “what it meant to her” – what difference her knowing that she had done it made to the value that she set on her life as a whole, or on some major part of it.<sup>215</sup>

In light of these quotes, it is plausible to think that, with regard to the philosopher in the above example, Adams might claim that we may count the accomplishment that she got for herself as something that she enjoys in her life, at least if she savors it.

If Adams were to make this claim, this would be similar to his claim concerning the woman who swims the channel – that is to say, this would be similar to his claim that “we may count her swimming the channel as something that she enjoys in her life, if she savors the achievement”. However, I myself do not think, in either of these two cases, that it is accurate to speak of *enjoyment*. Think of the matter this way. If someone is now swimming the channel, and if she is having an unpleasant, weary, cold, and anxious time, then we shouldn’t say that she is *enjoying* this swim. Further, we should not later say that she *enjoyed* this swim, since, after all, the truth is that she *didn’t* enjoy this swim. If we are going to use ‘enjoyment’ at all in this context, then what we should say is that, although this individual did not enjoy her swim in the channel, she does enjoy reflecting on her having swum the channel – indeed, reflecting on her accomplishment is something that she enjoys doing. And, naturally, all of this also holds, *mutatis mutandis*, for the philosophy case.

But, if what I have just said here is right, then what is the point that Adams is trying to make when he says that “we may count her swimming the channel as something that she enjoys in her life, if she savors the achievement”? I think that what Adams is unwittingly pushing toward here is a reliance not on the pro-attitude of enjoyment, but



rather on the pro-attitude of *cognitive endorsement*. This becomes especially clear when Adams says that “the value that she sets on it retrospectively seems, intuitively, quite relevant to how good it was for her that her life included this episode”. Adams seems to be pushing toward the claim that how prudentially valuable she *thinks* this episode is partly determines how prudentially valuable this episode in fact is. But, if this is indeed the claim that Adams is pushing here, then he isn’t really seizing on enjoyment anymore – rather, he is seizing on *cognitive endorsement*. If someone sets a value on something retrospectively, she isn’t employing her faculty of enjoyment – rather, she is making some sort of judgment as to how valuable the thing in question is, and to do this she must employ her faculty of judgment/belief. Since she is here setting a value on something, or judging/believing the thing to be valuable in some way or other, we can and should say that she is here employing her faculty of cognitive endorsement.

It is worth noting that Adams himself eventually admits that it probably isn’t true that A *must* enjoy X in order for X to count as an aspect of A’s welfare. Adams says:

It is hard to maintain without qualification that excellence is a constituent of our welfare only *as* enjoyed. Some aspect of the relevant excellence may not be accessible to one’s experience, and may therefore not affect one’s enjoyment. Real friendships and real accomplishments are more excellent than illusory ones, and are therefore better for us, even if the illusions would yield as much subjective enjoyment. Thus, as I claimed earlier, the excellence criterion keeps the enjoyment criterion from making my good too independent of what is external to me...I am not denying, of course, that real friendships and real achievements are enjoyed; so it does not follow from these examples that the excellence of what is not enjoyed *at all* can constitute part of one’s well-being. Probably it can, for it is plausible to think it remains better for oneself to do what is excellent when no available course of action affords any enjoyment. But a life rich only in that sort of excellence is no life to wish on a friend. It is in the enjoyment of excellence that a person’s good is primarily to be sought.<sup>216</sup>

In saying “probably it can” Adams means that it is likely that the excellence of what is not enjoyed *at all* can constitute part of one’s welfare.

Here (in section 32) I have argued that one need not enjoy X in order for X to be an aspect of one's welfare. It follows from this that all hybrid theories that rely on enjoyment fail. Thus we don't need to delve into the question of whether enjoyment is best combined with a relational or, instead, a non-relational sort of value.

### **33. Should we rely on cognitive endorsement?**

Even if we reject all hybrid theories that rely on enjoyment, that does not show that the no priority theory is in the right inasmuch as it relies on desire, for there are still other pro-attitudes to consider here. In particular, one might think that, instead of relying on desire, we ought to rely on cognitive endorsement.

Consider a hybrid theory that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is intrinsically good period and (2) A cognitively endorses X. There are different ways to think about the meaning of 'A cognitively endorses X'. This phrase might mean 'A judges/believes X to be intrinsically good period'. If this is the intended meaning, then we get a theory that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is intrinsically good period and (2) A judges/believes X to be intrinsically good period. This theory is implausible. It is intrinsically good period that various people somewhere far away from me – say, in India – are in good health right now. And I may – and, indeed, do – judge/believe that this is so. But it is not part of my welfare that various people in India are in good health right now. Now consider another possible meaning for 'A cognitively endorses X'. This phrase might mean 'A judges/believes X to be an aspect of his or her own welfare'. If this is the intended meaning, then we get a theory that says that X is an aspect of A's welfare if and

only if, and directly because, (1) X is intrinsically good period and (2) A judges/believes X to be an aspect of his or her own welfare. Is this hybrid theory plausible? No, it isn't.

In order to see why I say this, it may help if we first consider a simpler theory. Call this simpler theory *the belief welfare theory*. According to the belief welfare theory, X is an aspect of A's welfare if and only if, and directly because, A believes that X is an aspect of her welfare. We might say: 'The belief welfare theory is implausible. After all, it is circular. It explains well-being in terms of beliefs about well-being.' This may be true. But, whether it is or not, we should say more here. We should try to do a better job of pinning down exactly what is wrong with the belief welfare theory. Suppose that I am about to form (but haven't yet formed) the belief that X is an aspect of my welfare. And suppose that, when I do form this belief, this belief will be a *true* belief. Now let's ask: *Why* will I form this belief, that is, when I do in fact form it? The most natural answer here is this: I will somehow (accurately) perceive that X is an aspect of my welfare, and I will *thereby* form the belief that X is an aspect of my welfare. But notice that on the belief welfare theory this answer doesn't seem to make sense – on the belief welfare theory it doesn't make sense to say that I will somehow (accurately) perceive that X is an aspect of my welfare and *thereby* form the belief that X is an aspect of my welfare. I say this because on the belief welfare theory I can't somehow (accurately) perceive that X is an aspect of my welfare unless and until I have *already* formed the belief that X is an aspect of my welfare. After all, on the belief welfare theory my belief that X is an aspect of my welfare is what makes X be an aspect of my welfare in the first place.

The main point here, then, is this. On the belief welfare theory it is deeply mysterious as to what, exactly, is going on when we form our beliefs about our well-

being. This means, then, that on the belief welfare theory our beliefs about our well-being are deeply mysterious. However, the belief welfare theory explains our well-being in terms of our (deeply mysterious) beliefs about our well-being. So the belief welfare theory offers us a deeply mysterious explanation of our well-being. It is best to reject the belief welfare theory. This theory doesn't seem credible.

Now return to the hybrid theory that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is intrinsically good period and (2) A judges/believes X to be an aspect of his or her own welfare. This theory is going to be plagued by roughly the same problem that plagues the belief welfare theory. On this theory our beliefs about our own welfare are going to end up looking very mysterious. And in turn this theory is going to end up offering us a deeply mysterious explanation of our well-being. We should reject this theory. This theory doesn't seem credible.

Are there *any* plausible welfare theories that centrally rely on the pro-attitude of cognitive endorsement? Perhaps the most plausible theory of this sort is the one that says: X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) A cognitively endorses X. This hybrid theory is similar to the no priority theory. The only difference between this hybrid theory and the no priority theory is this: Whereas the no priority theory relies on desire, this hybrid theory relies on cognitive endorsement. Is this hybrid theory convincing? We need to know what 'A cognitively endorses X' means here. In order to avoid running into the same sort of problem that plagues the belief welfare theory, we should here steer clear of thinking of 'A cognitively endorses X' as meaning 'A judges/believes X to be an aspect of his or her

own welfare'. Here it seems best to think of 'A cognitively endorses X' as meaning 'A judges/believes X to be intrinsically good for herself in a relational, though non-prudential, way'. This leads us to a hybrid theory that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) A believes that X is good is intrinsically good for herself in a relational, though non-prudential, way. How convincing is this hybrid theory?

I am inclined to think that this is indeed the most plausible of all of the available hybrid theories that rely on cognitive endorsement. But, that being said, we should reject this theory. The main problem with this theory is that it is doubtful that, in all cases where X is an aspect of A's welfare, A has some belief to the effect that X is intrinsically good for herself (in any way at all). Here consider two points.

(1) Various things enter the well-being of babies. But I'm not sure that babies cognitively endorse much at all. It is part of their welfare to eat well and to sleep well. But do babies cognitively endorse these things? Do they ever judge that, or form the belief that, these things are intrinsically good for themselves (in any way)? I am inclined to doubt that they do, at least if they are very young (for instance, if they are newborns).

(2) There are cases where a belief to the effect that X is intrinsically good for oneself (in some way or other) seems to come *too late* – that is to say, there are various cases where X seems to be an aspect of A's welfare *prior* to A's ever forming a belief to the effect that X is intrinsically good for oneself (in some way or other). Sumner's bluegrass case seems to be such a case (see section 8): "Having never heard bluegrass, I chance on a band in the park and find that I like it."<sup>217</sup> Here the relationship of prudential

fit between A (on the one hand) and A's hearing bluegrass (on the other hand) seems to be in place even before A hears the bluegrass in the park and forms the belief that his hearing bluegrass is (in some way) intrinsically good for himself. In other words, just before A chances on the band in the park, it already seems to be true that A's hearing bluegrass is included within A's welfare as a *merely near-possible* aspect of it; and yet, at this point in time, A doesn't yet have any belief to the effect that hearing bluegrass is intrinsically good for himself.

The fundamental point for this section (section 33) is that it seems best for hybrid theorists to avoid relying on cognitive endorsement when they are formulating their theories. Perhaps the most plausible welfare theory that centrally relies on cognitive endorsement is the hybrid theory that we have just covered – that is, the one that says that X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) A believes that X is intrinsically good for herself in some relational, though non-prudential, way. Yet even this theory is dubious.

I should note that, aside from desire, enjoyment, and cognitive endorsement, there are other pro-attitudes that have been mentioned in the literature on well-being. For instance, Darwall sometimes speaks of the pro-attitude of *appreciative rapport*. Darwall says:

Experiences that make the greatest contributions to our welfare, I argue, are those we have when we are in appreciative rapport with forms of agent-neutral value – aesthetic values, for example or the value of sentient beings – whose significance is at least partly independent of their contribution to welfare.<sup>218</sup>

I don't know exactly what appreciative rapport is, but it seems to be something close to cognitive endorsement. If I am right about this, then I take it that hybrid theories that

centrally rely on appreciate rapport are going to be plagued by at least roughly the same problems that plague hybrid theories that centrally rely on cognitive endorsement.

Our findings in sections 31-33 seem to lead us to say the following: We should probably forget about enjoyment, cognitive endorsement, and appreciative rapport.

Desire seems to be the pro-attitude that the hybrid theorist should seize on.

### **34. How should a hybrid theory that relies on desire be formulated?**

Even given that desire is the pro-attitude that hybrid theorists should rely on, there is still the question of how, exactly, a hybrid theory that relies on desire ought to be formulated. We know how the no priority theory is formulated. It says that X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) X's obtaining satisfies some desire that A has (with the idea being that, if A does not desire X itself, then A must at least desire the basic good that X here instantiates for A). But there are other possibilities that we might consider here.

For example, a hybrid theorist might claim that X is an aspect of A's welfare if and only if, and directly because, (1) X is intrinsically good period and (2) A desires X. Is this welfare theory plausible? No, it isn't. I say this because there are cases where we desire things that are non-relationally and intrinsically good, where these desires of ours are fulfilled, and yet where it seems implausible to claim that these instances of desire-fulfillment advance our well-being. Here consider two cases that fit this description.

Case #1 is adapted from Derek Parfit's famous stranger case, which runs as follows: "Suppose that I meet a stranger who has what is believed to be a fatal disease.

My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured.”<sup>219</sup> Given the plausible assumption that the stranger’s being cured is non-relationally and intrinsically good, it follows that this case is a problem for any hybrid theory that relies on desire and non-relational goodness. I say this because it does not seem that this instance of desire-fulfillment really does advance Parfit’s welfare. Desires are fulfilled when the states that are their objects obtain. In this case Parfit’s desire is for the state *that the stranger is cured*. But, when this state obtains, its obtaining seems to be too independent of Parfit to advance his well-being.

Case #2 is one that I brought up in section 2. Suppose that a natural disaster (for instance, a tornado) strikes a town and victimizes many people there; and suppose that you form the desire that these people recover from their misfortune (maybe you even give a small donation). And now suppose that they do end up recovering. Their recovering seems not only to be good for them, but also to be intrinsically good period. Notice, though, that their recovering seems to be too independent of you to touch upon (or to enter) your welfare; and this seems especially to be true if you aren’t even aware of their recovering.

The upshot of our reflection on these two cases is that, if a hybrid theorist is going to rely on desire, then she shouldn’t rely on any sort of non-relational goodness – rather, she should rely on some sort of relational goodness. Since relational values are always for agents, it seems that there is no serious worry about relational values being too independent of agents to touch upon (or to enter) their own welfare.<sup>220</sup> Naturally, in formulating the no priority theory I have made sure to rely on the desire-cum-relational



goodness combination. This combination seems to be much better than the desire-cum-non-relational goodness combination.

Even if we have rightly made up our minds to rely on the desire-cum-relational goodness combination, there are still two options that we need to decide between. It is open to us to go with the following formulation: X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) A desires X. Or, alternatively, we might go with what the no priority theory says: X is an aspect of A's welfare if and only if, and directly because, (1) X is, for A, an instance of one of the basic goods (that is, where the basic goods are not being conceived of as prudential goods) and (2) X's obtaining satisfies some desire that A has (with the idea being that, if A does not desire X itself, then A must at least desire the basic good that X here instantiates for A).

We should go with what the no priority theory says. The problem with the first formulation is that it seems that there are cases where X is an aspect of one's welfare (that is, where X is a *merely near-possible* aspect of one's welfare), and yet where one lacks the desire for X – either because one doesn't know about X or for some other reason.

Here we might think of the case of the young academic. There is a period of time during which she believes that her book is a failure, and (in turn) during which she has lost the desire to finish her book. Yet finishing her book seems to be included within her welfare at this time (as a *merely near-possible* aspect of it). Though the first formulation from above can't handle this case, the no priority theory can handle this case. After all,

the no priority theory can say (1) that the young academic desires accomplishment, and (2) that finishing the book satisfies the young academic's desire for accomplishment.

Similar remarks apply to the bluegrass case. There is a point in time where A's listening to bluegrass is a merely near-possible aspect of A's welfare, even though A hasn't yet formed any desire for listening to bluegrass. The best way to handle cases such as this one is simply to abandon the requirement that A must desire X in order for X to be an aspect of A's welfare. What we should say is that X can't be an aspect of A's welfare unless X's obtaining satisfies a desire that A has. Assuming that A has the desire for aesthetic experience, and assuming that A's listening to bluegrass satisfies A's desire for aesthetic experience, we can say that A's hearing bluegrass satisfies a desire that A has – namely, the desire for aesthetic experience. This is how the no priority theory handles cases such as this one; and this seems to be the best way to handle cases such as this one.

### **35. Closing comments**

If the above remarks are on target, then it seems reasonable to think that the no priority theory is the best available hybrid theory. There is, however, one final question to consider here. That is the question of how, exactly, the no priority theory ought to be defended. It is important that the no priority theory be defended in the particular way that I have defended it – that is, by way of emphasizing that human desire with respect to the basic goods is remarkably widespread and stable.

Think of the point this way. If A lacks the desire for accomplishment, or for friendship, or for aesthetic experience, etc., then, according to the no priority theory, this

non-desired good isn't part of A's welfare. OL theorists find this implication of the no priority theory unacceptable. However, I have argued that there is no genuine worry for the no priority theory here, because (1) the vast majority of humans just do desire accomplishment, friendship, aesthetic experience, etc., and stably so, and because (2) in those very rare cases where the desires for these goods are missing, we don't have sufficient reasons for thinking that the relevant non-desired goods really are part of the welfare of the agents in question.

It is true, of course, that someone might accept the no priority theory and, in addition, reject the claim that human desire with respect to the basic goods is remarkably widespread and stable. That is to say, someone might accept the no priority theory and, in addition, allow that people's desires for the basic goods are unstable, subject to flux, etc. But I can't help but think that this is a terrible position to take. It does not seem at all right to say that friendship is sometimes part of my welfare, and sometimes not. The truth seems to be that friendship is a stable part of my welfare. In order for the no priority theory to entail the truth here – that is, in order for the no priority theory to entail that friendship is a stable component of my welfare – it must be true that I stably desire friendship. I have argued that this is true. And, of course, the argument doesn't just concern me and my stable desire for friendship. The argument concerns all humans and their stable desires for all of the basic goods (excepting, of course, for some very rare cases, such as the case of friendship-less autistics and their not desiring friendship).

To end: We need a theory of human well-being that pays due respect to both the *good* and the *for* parts of 'good for' (that is, of '*prudentially* good for'). The no priority theory gives us precisely what we need. It is the best theory of human welfare.

## Bibliography

- Adams, Robert. (1999) *Finite and Infinite Goods* (New York: Oxford University Press).
- Augustine. (1961) *Confessions* (New York: Penguin Books).
- Baumeister, Roy. (1989) *Masochism and the Self* (Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers).
- Bowers and Andersen. (2007) "Cognitive-Behavior Therapy with Eating Disorders: The Role of Medications in Treatment," *Journal of Cognitive Psychotherapy: An International Quarterly* 21 (1), 16-27.
- Brandt, Richard. (1998) *A Theory of the Good and the Right* (Amherst, New York: Prometheus Books).
- Brink, David. (1989) *Moral Realism and the Foundations of Ethics* (New York: Cambridge University Press).
- Brink, David. (2006) "Some Forms and Limits of Consequentialism." In *The Oxford Handbook of Ethical Theory*. Ed. David Copp (New York: Oxford University Press).
- Chan, David. (2004) "Are There Extrinsic Desires?," *Nous* 38 (2), 326-350.
- Crisp, Roger. (2006) *Reasons and the Good* (Oxford: Clarendon Press).
- Darwall, Stephen. (1992) "Internalism and Agency." In *Philosophical Perspectives, 6: Ethics, 1992*. Ed. James Tomberlin (Atascadero, CA: Ridgeview Publishing Company).
- Darwall, Stephen. (2002) *Welfare and Rational Care* (Princeton: Princeton University Press).
- Darwall, Stephen. (2006) "Precis of *Welfare and Rational Care*," *Philosophical Studies* 130, 579-584.
- Doctor and Kahn. (2000) *The Encyclopedia of Phobias, Fears, and Anxieties* (New York: Facts on File, Inc.).
- Donnellan, Leary, and Robledo. (2006) "I Can't Get Started: Stress and the Role of Movement Differences in People with Autism." In *Stress and Coping in Autism*. Eds. Baron, J. Groden, G. Groden, and Lipsitt (New York: Oxford University Press).
- Dworkin, Ronald. (2000) *Sovereign Virtue* (Cambridge, MA: Harvard University Press).
- Edens, Lilienfeld, Marcus, and Poythress Jr. (2006) "Psychopathic, Not Psychopath: Taxometric Evidence for the Dimensional Structure of Psychopathy," *Journal of Abnormal Psychology* 115 (1), 131-144.
- Feldman, Fred. (2004) *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism* (Oxford: Clarendon Press).
- Finnis, John. (1980) *Natural Law and Natural Rights* (Oxford: Clarendon Press).
- Finnis, John. (1998) *Aquinas* (New York: Oxford University Press).
- Gilbert, Daniel. (2005) *Stumbling on Happiness* (New York: Vintage Books).
- Gilbert, Paul. (2001) *Overcoming Depression: A Step-by-Step Approach to Gaining Control Over Depression* (New York: Oxford University Press).

- Gomez-Lobo, Alfonso. (2002) *Morality and the Human Goods: An Introduction to Natural Law Ethics* (Washington, DC: Georgetown University Press).
- Grandin, Temple. (2006) "Stopping the Constant Stress." In *Stress and Coping in Autism*. Eds. Baron, J. Groden, G. Groden, and Lipsitt (New York: Oxford University Press).
- Griffin, James. (1986) *Well-Being: Its Meaning, Measurement and Moral Importance* (Oxford: Clarendon Press).
- Griffin, James. (1996) *Value Judgment: Improving Our Ethical Beliefs* (Oxford: Clarendon Press).
- Guay, Knight, Ruscio, and Hare. (2007) "A Taxometric Analysis of the Latent Structure of Psychopathy: Evidence for Dimensionality," *Journal of Abnormal Psychology* 116 (4), 701-716.
- Hare, Robert. (1970) *Psychopathy: Theory and Research* (New York: John Wiley and Sons, Inc.).
- Heathwood, Chris. (2005) "The problem of defective desires," *Australasian Journal of Philosophy* 83 (4), 487-504.
- Hittinger, Russell. (1987) *A Critique of the New Natural Law Theory* (Notre Dame, IN: University of Notre Dame Press).
- Hobson, Peter. (2007) "On Being Moved in Thought and Feeling: An Approach to Autism." In *New Developments in Autism: The Future is Today*. Eds. Perez, Gonzalez, Comi, and Nieto (Philadelphia: Jessica Kingsley Publishers).
- Hollon, Steven. (2006) "Cognitive Therapy in the Treatment and Prevention of Depression." In *The Interpersonal, Cognitive, and Social Nature of Depression*. Eds. Joiner, Brown, and Kistner (Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers).
- Hooker, Brad. (1996) "Does Moral Virtue Constitute a Benefit to the Agent?" In *How Should One Live?* Ed. Roger Crisp (New York: Oxford University Press).
- Hooker, Brad. (2000) *Ideal Code, Real World* (New York: Oxford University Press).
- Hurka, Thomas. (1987) "'Good' and 'Good For,'" *Mind* 96 (381), 71-73.
- Hurka, Thomas. (1993) *Perfectionism* (New York: Oxford University Press).
- Kagan, Shelly. (1998) *Normative Ethics* (Boulder, Colorado: Westview Press).
- Kinoy, Holman, and Lemberg. (1999) "The Eating Disorders: An Introduction." In *Eating Disorders: A Reference Sourcebook*. Eds. Lemberg and Cohn (Phoenix, Arizona: Oryx Press).
- Korsgaard, Christine. (2004) "The Myth of Egoism." In *Practical Conflicts: New Philosophical Essays*. Eds. Baumann and Betzler (New York: Cambridge University Press).
- Kraut, Richard. (2007) *What Is Good and Why: The Ethics of Well-Being* (Cambridge, MA: Harvard University Press).
- Lewis, David. (1988) "Desire as Belief," *Mind* 97, 323-332.
- Lewis, David. (1989) "Dispositional Theories of Value." In *The Aristotelian Society Supplementary Volume LXIII*.
- Lewis, David. (1996) "Desire as Belief II," *Mind* 105, 303-313.

- Martens, Wilhelm. (2002) "The Hidden Suffering of the Psychopath," *Psychiatric Times* 19 (1) (<http://www.psychiatrictimes.com/p020128.html>).
- Martens, Wilhelm and George Palermo. (2005) "Loneliness and Associated Violent Antisocial Behavior: Analysis of the Case Reports of Jeffrey Dahmer and Dennis Nilsen," *International Journal of Offender Therapy and Comparative Criminology* 49 (3), 298-307.
- Medline Plus Medical Encyclopedia (from the National Institutes of Health). (2008) "Antisocial Personality Disorder," (<http://www.nlm.nih.gov/medlineplus/ency/article/000921.htm>).
- Mickley, Diane. (1999) "Medical Dangers of Anorexia Nervosa and Bulimia Nervosa." In *Eating Disorders: A Reference Sourcebook*. Eds. Lemberg and Cohn (Phoenix, Arizona: Oryx Press).
- Mill, J. S. (1998) *Utilitarianism* (New York: Oxford University Press).
- Moore, Andrew and Roger Crisp. (1996) "Welfarism in Moral Theory," *Australasian Journal of Philosophy* 74 (4), 598-613.
- Moore, G. E. (1951) *Principia Ethica* (Cambridge, Great Britain: Cambridge University Press).
- Murphy, Mark. (1999) "The Simple Desire-Fulfillment Theory," *Nous* 33 (2), 247-272.
- Murphy, Mark. (2001) *Natural Law and Practical Rationality* (New York: Cambridge University Press).
- Ochberg, Brantley, Hare, Houk, Ianni, James, O'Toole, and Saathoff. (2004) "Lethal Predators: Psychotic, Sadistic, and Sane." In *Profilers*. Eds. Campbell and DeNevi (Amherst, New York: Prometheus Books).
- Parfit, Derek. (1984) *Reasons and Persons* (New York: Oxford University Press).
- Peterson and Mitchell. (1999) "Cognitive-Behavioral Therapy and Eating Disorders." In *Eating Disorders: A Reference Sourcebook*. Eds. Lemberg and Cohn (Phoenix, Arizona: Oryx Press).
- Quinn, Warren. (1993) "Putting Rationality in Its Place." In *Value, Welfare, and Morality*. Eds. Frey and Morris (New York: Cambridge University Press).
- Railton, Peter. (1989) "Naturalism and Prescriptivity," *Social Philosophy and Policy* 7 (1), 151-174.
- Railton, Peter. (2003) *Facts, Values, and Norms* (New York: Cambridge University Press).
- Rapoport, Judith. (1989) *The Boy Who Couldn't Stop Washing: The Experience and Treatment of Obsessive-Compulsive Disorder* (New York: E. P. Dutton).
- Rawls, John. (1999) *A Theory of Justice* (Cambridge, MA: The Belknap Press of Harvard University Press).
- Raz, Joseph. (2004) "The Role of Well-Being," *Philosophical Perspectives* 18: *Ethics*, 2004, 269-294.
- Regan, Donald. (2004) "Why Am I My Brother's Keeper?" In *Reason and Value*. Eds. Wallace, Pettit, Sheffler, and Smith (Oxford: Oxford University Press).
- Rosati, Connie. (1995) "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105, 296-325.
- Rosati, Connie. (1996) "Internalism and the Good for a Person," *Ethics* 106, 297-326.

- Rosati, Connie. (2000) "Brandt's Notion of Therapeutic Agency," *Ethics* 110, 780-811.
- Rosati, Connie. (2006a) "Darwall on Welfare and Rational Care," *Philosophical Studies* 130, 619-635.
- Rosati, Connie. (2006b) "Personal Good." In *Metaethics after Moore*. Eds. Horgan and Timmons (Oxford: Clarendon Press).
- Rosati, Connie. (2006c) "Preference Formation and Personal Good." In *Preferences and Well-Being*. Ed. Serena Olsaretti (New York: Cambridge University Press).
- Schreibman, Laura. (2005) *The Science and Fiction of Autism* (Cambridge, Massachusetts: Harvard University Press).
- Schroeder, Timothy. (2004) *Three Faces of Desire* (New York: Oxford University Press).
- Sidgwick, Henry. (1981) *The Methods of Ethics* (Indianapolis: Hackett Publishing Company).
- Smith, Michael. (1994) *The Moral Problem* (Malden, MA: Blackwell Publishing).
- Sobel, David and David Copp. (2001) "Against Direction of Fit Accounts of Belief and Desire," *Analysis* 61, 44-53.
- Sobel, David. (2009) "Subjectivism and Idealization," *Ethics* 119, 336-352.
- Solomon and Haaga. (2004) "Cognitive theory and therapy of depression." In *Cognitive Therapy Across the Lifespan: Evidence and Practice*. Eds. Reinecke and Clark (Cambridge, UK: Cambridge University Press).
- Stampe, Dennis. (1987) "The Authority of Desire," *The Philosophical Review* 96 (3), 335-381.
- Stocker, Michael. (1979) "Desiring the Bad: An Essay in Moral Psychology," *The Journal of Philosophy* 76 (12), 738-753.
- Sumner, L.W. (1996) *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press).
- Sumner, L.W. (2000) "Something in Between." In *Well-Being and Morality: Essays in Honor of James Griffin*. Eds. Crisp and Hooker (Oxford: Clarendon Press).
- Thurstin, Adrian. (1999) "Behavioral, Physical, and Psychological Symptoms of Eating Disorders." In *Eating Disorders: A Reference Sourcebook*. Eds. Lemberg and Cohn (Phoenix, Arizona: Oryx Press).
- Velleman, J. David. (1992) "The Guise of the Good," *Nous* 26 (1), 3-26.
- Wagner, Susan. (1999) "Eating Disorder Treatment Stories: Four Cases." In *Eating Disorders: A Reference Sourcebook*. Eds. Lemberg and Cohn (Phoenix, Arizona: Oryx Press).
- Wall, Steven. (2007) "Perfectionism in Moral and Political Philosophy," *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/perfectionism-moral/>).
- Williams, Bernard. (1981) *Moral Luck* (New York: Cambridge University Press).
- Williams, Donna. (1994) *Somebody Somewhere: Breaking Free from the World of Autism* (New York: Times Books).
- Wolterstorff, Nicholas. (2008) *Justice: Rights and Wrongs* (Princeton, NJ: Princeton University Press).

---

<sup>1</sup> Sometimes DF theories are qualified in certain ways. For instance, sometimes DF theories say that X is an aspect of A's welfare if and only if, and directly because, (1) A desires X and (2) X's obtaining is something that A experiences (see Griffin 1986, 13).

<sup>2</sup> OL theorists may not see this as platitudinous. But here I am just trying to explain the way in which DF theorists typically criticize OL theories. DF theorists do think about welfare as a very *personal* sort of value; and DF theorists do think that OL theories fail to capture the distinctively personal nature of welfare.

<sup>3</sup> Connie Rosati writes about prudential internalism (see Rosati 1996, 297-326), and so does Peter Railton (see Railton 2003, 47).

<sup>4</sup> In coming up with this case, I am drawing from Parfit's stranger-on-the-train case (see Parfit 1984, 494).

<sup>5</sup> Brad Hooker is an OL theorist who levels (roughly) this objection against hybrid theories (he calls them mixed theories). He says: "From the point of view of [objective] list theorists, [the] mixed view is mistaken to claim that a state of affairs can constitute a benefit to you only if you endorse it. For example, some achievement in your life might constitute at least a small benefit to you, might have contributed at least some small meaning to your life, even if you never cared about it" (see Hooker 2000, 41).

<sup>6</sup> See Adams 1999, 83-101; see Darwall 2002, 73-104; and see Rosati 2006b, 107-131 and Rosati 2006c, 33-64. For the quote from Rosati, see Rosati 2006c, 63, n. 45.

<sup>7</sup> For Schroeder's discussion of (and critique of) the standard theory, see Schroeder 2004, 10-37.

<sup>8</sup> See Schroeder 2004, 11.

<sup>9</sup> See Schroeder 2004, 11.

<sup>10</sup> See Schroeder 2004, 15.

<sup>11</sup> Michael Smith has gone into some detail in discussing the claim that beliefs and desires have different directions of fit (see Smith 1994, 111-116).

<sup>12</sup> Smith expresses the worry that different directions of fit talk might be overly metaphorical (see Smith 1994, 112). With respect to this issue, also see Sobel and Copp 2001, 44-53.

<sup>13</sup> See Schroeder 2004, 155.

<sup>14</sup> In his article "The Guise of the Good" David Velleman has argued that the case of perverse desire provides us with a counterexample to the thesis that one can't desire X unless one believes that X is (somehow) good. Velleman states: "[W]hat makes a desire perverse is that its propositional object implies that it is inappropriate. That is, the perverse subject desires that something undesirable occur, and its being undesirable is part of the description under which he desires it" (see Velleman 1992, 18).

<sup>15</sup> Am I being too quick in accepting the standard theory of desire? The standard theory says that one of my mental states is a desire if and only if, and because, it disposes me to bring about its object. It seems clear that, in order for one of my mental states to be a desire, it must dispose me to bring about its object. But it might be thought, with a fair amount of plausibility, that I should leave open the possibility that a desire is more than *just* a disposition to bring about its object. But what, besides being a disposition, could a desire be? Here consider the following remarks. There should be a way of distinguishing desires (that is, intrinsic desires) from merely instrumental desires (if there are such things), intentions, and also perhaps habits. Here we might try appealing to the belief thesis – desires essentially involve beliefs to the effect that their objects are (in some way) good, whereas instrumental desires, intentions, and habits do not essentially involve beliefs to the effect that their objects are (in some way) good. But, as I have indicated, the belief thesis should probably be shed – one can probably desire X without believing that X is good in some way.



---

The best way to distinguish desires from instrumental desires and intentions is to appeal to the notion of intrinsic motivational force. Desires are essentially intrinsic motivations, whereas instrumental desires and intentions are not essentially intrinsic motivations. One can instrumentally desire or intend something that one isn't intrinsically motivated to bring about – not so with desire, though. The harder distinction to draw here is the one between desires and habits. Intuitively, habits are essentially intrinsic motivations, and, intuitively, you can do something out of habit and yet not desire to do it (for a discussion of this issue, see Schroeder 2004, 22). Here I see two options worth considering. One is to say (1) that desires = intrinsic motivations and (2) that, since habits are essentially intrinsic motivations, it follows that all habits are desires. On this option, then, we are going against common sense inasmuch as we are saying that all habits are desires. But perhaps that is fine, all things considered, since often the best theory of something has at least one implication that runs contrary to common sense. The other option is to find some way of making it clear that, although habits are essentially intrinsic motivations, it doesn't follow that all habits are desires. The only way I can think of to make this option work out is to say that, although habits are essentially intrinsic motivations, habits don't essentially involve reward signals in the brain, whereas desires are essentially intrinsic motivations and essentially involve reward signals in the brain. Here we would need to draw on some of Schroeder's work on the relationship between desire and reward (see Schroeder 2004). I do not know, though, whether this option is, in the end, a workable one. At any rate, my official position is that I accept the standard theory of desire (which says that desires are mental states that are *just* dispositions to bring about their objects) – though I do admit that I am at least somewhat open to being persuaded that there is more to one of my mental state's being a desire than just its disposing me to bring about its object.

<sup>16</sup> See Lewis 1989, 113-137.

<sup>17</sup> See Lewis 1988, 323-332, and see Lewis 1996, 303-313.

<sup>18</sup> For a discussion of this matter, see Griffin 1996, 19-21. Griffin discusses the taste model, according to which things are valuable because they are desired; and Griffin also discusses the perception model, according to which things are desired because they are seen to be valuable.

<sup>19</sup> For a discussion of this issue, see Griffin 1996, 19-36.

<sup>20</sup> See Quinn 1993, 32.

<sup>21</sup> See Murphy 1999, 253.

<sup>22</sup> See Brink 1989, 225-226.

<sup>23</sup> Alan could, of course, have the belief that baklava is valuable in that it is tasty, but Brink's point here seems to be that we don't *have* to posit any such belief on Alan's part in order for it to be true that Alan desires baklava – the desire could just be brute (or spontaneous).

<sup>24</sup> In the introduction to his book, Schroeder says that his book is about intrinsic desires, except for some occasional, passing remarks on instrumental desires (see Schroeder 2004, 5). So, in discussing the standard theory of desire in Chapter 1 of his book, it is clear that Schroeder is presenting the standard theory of desire as a theory of what intrinsic desires are.

<sup>25</sup> For instance, see Chan 2004, 326-350.

<sup>26</sup> See Murphy 1999, 253.

<sup>27</sup> See Murphy 1999, 253.

<sup>28</sup> See Murphy 1999, 253-254.

---

<sup>29</sup> I am assuming that dispositions by their very nature have motivational force in their own right – that is, I am assuming that there are no purely instrumental dispositions. Thus I am using the word ‘disposition’ in a way that is stronger than many others would use this word. Someone might say: ‘I am disposed to be snappish with my wife when hungry. But I do not desire to be snappish with my wife when hungry. This shows that, whatever a desire is, it can’t *just* be a disposition.’ At least if this is any ordinary case, I would say that this man *isn’t* disposed to be snappish with his wife when hungry. I would say that this man’s motivational force is flowing not to the state *that he be snappish with his wife when hungry*, but rather to, say, the state *that he be left in peace when he is irritable*. And I would say that, although this man often pursues the state *that he be snappish with his wife when hungry*, he often pursues this state not because he is disposed to bring *it* about, but rather because (1) he is disposed to bring about the state *that he be left in peace when he is irritable* and (2) he believes that being snappish with his wife is an effective way of bringing about this latter state (say, because he believes that, by being snappish with his wife, he will get her to stop nagging him). My general point here is that, in the context of discussions of a what a desire is, I use the word ‘disposition’ in a strong way – in particular, I think of dispositions and intrinsic motivational force as going hand-in-hand.

<sup>30</sup> As I mentioned above (see n. 1), DF theories are sometimes qualified in certain ways.

<sup>31</sup> See Lewis 1989, 113-137, and see Railton 2003, 43-68. Also, see Railton 1989, 151-174.

<sup>32</sup> For Brink’s discussion of his objectivism about well-being, see Brink 1989, 217-236. Brink discusses his naturalism in various places – for example, see Brink 1989, 238-240.

<sup>33</sup> Sumner lays out his theory of welfare in Sumner 1996, 138-183.

<sup>34</sup> See Railton 2003, 47.

<sup>35</sup> See Brandt 1998, 249-250.

<sup>36</sup> See Brandt 1998, 250. (Incidentally, I should note that Brandt traces the example involving the religious skeptic back to Parfit and Griffin – see Brandt 1998, 250, n. 1.)

<sup>37</sup> See Brandt 1998, 30.

<sup>38</sup> See Sumner 1996, 133.

<sup>39</sup> See Schroeder 2004, 191, n. 9.

<sup>40</sup> See Mill 1998, 83.

<sup>41</sup> See Rapoport 1989, 2.

<sup>42</sup> See Rapoport 1989, 87-88.

<sup>43</sup> For example, many autistics engage in self-injurious behavior, but there are debates about what is motivating this behavior (see Schreibman 2005, 140). Also, many autistics engage in echolalia (or the speech act of parroting what others say). But there are also debates about what motivates this sort of behavior (see Schreibman, 243-244).

<sup>44</sup> See Railton 2003, 54.

<sup>45</sup> I should thank Alex Pruss for this point.

<sup>46</sup> See Griffin 1996, 21.

---

<sup>47</sup> See Heathwood 2005, 490.

<sup>48</sup> I say ‘his DF theory’, but I should stress that Heathwood doesn’t endorse the DF theory that he here defends. Heathwood says: “The actualist theory defended here is definitely not unfalsifiable. In fact, I myself think it is probably false! But this is due to a different objection. My point here has not been to endorse once and for all the particular actualism presented here. My point rather is to show that the problem of defective desires is, after all, no problem for even very simple forms of actualism...” (see Heathwood 2005, 500).

<sup>49</sup> See Heathwood 2005, 493.

<sup>50</sup> My own stance on this issue is that, while it is too strong to say that A must desire X in order for X to be a merely near-possible aspect of A’s welfare, it is quite right to think that X must hook into A’s actual motivational structure/desire-set in order for X to be a merely near-possible aspect of A’s welfare. What we should hold is that X’s obtaining must satisfy a desire that A actually has in order for X to be a merely near-possible aspect of A’s welfare – see section 2 and section 22. If we hold this position, then we can say to A: ‘This aspect of your welfare, whatever it may be, is *for you* in that it is positively linked to your actual motivational structure – it is *for you* in that its obtaining satisfies a desire that you actually have.’

<sup>51</sup> See Railton 2003, 49.

<sup>52</sup> See Railton 2003, 52 and 54.

<sup>53</sup> See Lewis 1989, 113.

<sup>54</sup> See Lewis 1989, 114-115.

<sup>55</sup> See Lewis 1989, 115.

<sup>56</sup> See Smith 1994, 146.

<sup>57</sup> With respect to Brandt’s welfare hedonism, see Brandt 1998, 246-247; and, with respect to Railton’s being influenced by Brandt, see Railton 2003, 36, n. 14.

<sup>58</sup> See Brandt 1998, 113.

<sup>59</sup> See Brandt 1998, 113.

<sup>60</sup> See Brandt 1998, 13.

<sup>61</sup> See Brandt 1998, 12.

<sup>62</sup> This quote is from the foreword to Brandt’s *A Theory of the Good and the Right* (see Brandt 1998). Singer is the author of this foreword. (See v-ix for the foreword, and see viii for this particular quote.)

<sup>63</sup> See Rawls 1999, 365-372; see Railton 2003, 43-68; and see Lewis 1989, 113-137.

<sup>64</sup> Again, see Railton 2003, 43-68.

<sup>65</sup> For a discussion of something close to full information (without being quite full), see Brandt 1998, 12-13. I should note, though, that this discussion takes place not in the context of Brandt’s theory of well-being, but rather in the context of Brandt’s theory of rational desire. (Again, Brandt isn’t a hypothetical DF theorist – rather, he is a welfare hedonist.)

<sup>66</sup> See Lewis 1989, 113-137.

---

<sup>67</sup> See Lewis 1989, 121.

<sup>68</sup> See Lewis 1989, 125.

<sup>69</sup> See D. Gilbert 2005, 125-129.

<sup>70</sup> See D. Gilbert 2005, 125.

<sup>71</sup> See D. Gilbert 2005, 127.

<sup>72</sup> See D. Gilbert 2005, 138-139.

<sup>73</sup> See D. Gilbert 2005, 138.

<sup>74</sup> See Brink 2006, 390-391.

<sup>75</sup> See Lewis 1996, 305.

<sup>76</sup> See Brandt 1998, 113, where Brandt says: “One implication of our definitions [of rational and irrational desires] may be surprising. It arises from the fact that some valences, or dispositions to enjoy something, may resist extinction by inhibition and anything else, since they have been so firmly learned at an early age. By my definition these qualify as rational. For I use ‘rational’ as the contradictory of ‘irrational’ and have defined an ‘irrational’ desire (etc.) as one that would extinguish after cognitive psychotherapy. If a desire will not extinguish, then it is not irrational. This result is consistent with the general view that a desire (etc.) is rational if it has been influenced by facts and logic as much as possible. Unextinguishable desires meet this condition.”

<sup>77</sup> See Sobel 2009, 343.

<sup>78</sup> See Sobel 2009, 344.

<sup>79</sup> See Sobel 2009, 343.

<sup>80</sup> Railton seems to think that being fully informed entails that one is provided with facts about how things would go for one if such-and-such were to happen. If you read what Railton says about “the earlier Beth” and “the later Beth” in his article “Facts and Values”, then this comes out fairly clearly. For instance, Railton says: “Yet it is the later Beth who feels this force, for it is she who has the information. What force would the views of the later Beth have for the earlier Beth, were they somehow to become known to her?” (see Railton 2003, 51). With respect to the earlier Beth (who is an accountant), the question is that of what it would be like for her if she were to quit her accountant job and take up writing full time. The later Beth has actually done this and found out, much to her dismay, that she is not fit for writing full time. The earlier Beth does not know that she is not fit for writing full time, but, were this information to be given to her, this might cause her to lose her desire to become a full time writer (or it might at least cause her to desire that she not desire to become a full time writer). That, roughly put, is what Railton is thinking here. However, even if Railton is right that there is a determinate fact of the matter about what it would be like for the earlier Beth if she were to become a writer full time, this is just one case. Railton needs the point to generalize here, but, really, there doesn’t seem to be any general truth to the effect that there are facts about what it would be like to have various ways one’s life might go be actual. For there are too many cases – the case with Dennis is one of them – where there is no set fact about what it would be like for oneself if one’s life were to go in such-and-such a direction, or if such-and-such a state of affairs were to obtain.

<sup>81</sup> There is, perhaps, a way out for hypothetical DF theorists here – but this way out requires the assumption of determinism. Suppose that, as Dennis is being fully informed at t1, he is told that determinism is true, that the future is therefore completely settled, and that it is a set fact that at t2 he will be working at the firm and will be miserable there. If Dennis were told this while being fully informed at t1, then he probably

---

would lose his desire to work at the firm. So this probably would give hypothetical DF theorists the result that they are looking for here. However, two things should be said about all of this. First, there seems to be something illegitimate going on if hypothetical DF theorists go ahead and simply build determinism into their theories of well-being. To my knowledge, there are no hypothetical DF theorists who have made this move. It is true, I suspect, that most hypothetical DF theorists are determinists. But my point here is that (to my knowledge) none of them go so far as to build determinism into their accounts of well-being. The reason for this, I suspect, is that they know that there is something illegitimate about this move. At the very least, this move is seriously controversial and thus very costly. Second, it is worth noting that it is hard to know exactly what would happen to someone's psychology if it were made clear to him that determinism is true. Many of us (and I include myself here) see determinism as being a very real threat to human freedom and to the possibility of living a deeply meaningful human life. So, for many of us, it seems that, if it were made clear to us that determinism is true and that the future is therefore completely settled, this may well cause us to fall into some sort of nihilistic and/or paralytic state. Applying this point to Dennis, we can say the following: If determinism is assumed to be true, and if (as he is being fully informed at t1) he is told that determinism is true, that the future is therefore totally settled, and that it is a set fact that at t2 he will be working at the firm and will be miserable there, then he may fall into a nihilistic and/or paralytic sort of state – in which case his beliefs and desires may not function in anything like a predictable or well-behaved way (and thus in which case, strangely enough, Dennis may retain his desire to work at the firm, that is, even in spite of his having been told that he will work there and will be miserable there).

<sup>82</sup> Hypothetical DF theorists may retort: 'True, before Dennis works at the firm, there is no set fact as to what things would be like for him if he were to work there. But there *are* set facts of the following sort: If Dennis were to work at the firm, and if (while working there) he would be focusing on cases of such-and-such a sort, working seventy or more hours in a week, dealing with a super-cranky boss, etc., then he would be miserable. When Dennis is being fully informed, he will be told these set facts. And, once he is told these set facts, it stands to reason that he will no longer desire to work at the law firm (full stop) – rather, what he will desire is to work at the law firm, provided that such-and-such conditions obtain (say, that he is working on cases of a certain type, that his hours aren't insane, that his boss is relatively easy to deal with, etc.). The point, then, is that Dennis will refine, or fine-grain, his desires in light of full information, and, once he does this, Dead Sea apple cases won't be much of a problem for hypothetical DF theories.' This response deserves to be taken seriously, but ultimately it is unconvincing. Since people do not typically fine grain their desires in this way in real life, it is doubtful that people would fine-grain their desires in this way if they were in the hypothetical setting. Consider Dennis. It is not as though Dennis is being told all that much in the hypothetical setting that he wasn't already aware of (in real life). True, in the hypothetical setting he may learn some facts about the job that he didn't previously know. But still, when he first formed his desire to work at the firm in real life, he already knew that there were various possible downsides to the job – yet he formed the desire to work at the firm, full stop, anyway (he didn't form the desire to work at the firm, provided that such-and-such conditions obtain). And this, I think, is just a general fact about how our desires typically operate: They do not typically operate in a fine-grained way. Suppose (as is actually the case) that I am applying for philosophy jobs. For each job that I apply for, I am well aware that there are various possible downsides (say, because the location is bad, or because the teaching load is super-high, or whatever). In spite of my knowledge of these various possible downsides, I do not fine-grain my desires such that I desire to get this or that job, provided that certain conditions obtain (conditions, that is, wherein these downsides are either eliminated or mitigated). Rather, what I do is simply form the desire to get this or that job (full stop). As for *why* I do this, I don't know. Perhaps it has something to do with a human need to keep complex things as simple as possible: We form coarse-grained evaluative beliefs (say, the belief that this job is a good job) and coarse-grained desires (say, the desire to get this job) because that helps us keep our complex lives tolerably simple. Regardless of what the true explanation (for why we do not typically fine-grain our desires) is, the point is just that it is a fact about humans that we do not typically fine-grain our desires. Thus I don't think that the fine-graining of desire maneuver is ultimately going to help hypothetical DF theorists all that much when it comes to Dead Sea apple cases. (This maneuver may help to *some* extent – but the fundamental problem will still be there.)

<sup>83</sup> Railton is explicit that the non-belief properties of an individual are to be held constant (or as near as possible to constant) as she is being fully and vividly informed in the hypothetical setting. In particular, I

---

here have in mind Railton's saying: "[I]nsofar as possible, the idealization holds fixed the individual's non-belief properties, so that the contribution of these features to desire-formation would remain largely the same" (see Railton 2003, 58).

<sup>84</sup> See Rosati 1995, 296-325.

<sup>85</sup> Again, see Rosati 1995, 296-325.

<sup>86</sup> See Railton 2003, 36-37, n. 15.

<sup>87</sup> See Railton 2003, 54.

<sup>88</sup> See Railton 2003, 54.

<sup>89</sup> It seems that Railton is here assuming that HA and A will remain identical in terms of non-belief properties/personality. In other words, the only difference that Railton seems to have in mind here is a difference in (value-free factual) knowledge between HA and A.

<sup>90</sup> See Rosati 1995, 313.

<sup>91</sup> I think that the arguments against hypothetical DF theories that I have made here in Chapter 1 show that, even when hypothetical DF theories are offered as *revisionist* accounts of well-being, they still fail. On a revisionist account the aim isn't to capture exactly what we mean by 'well-being' in ordinary language – rather, the aim is merely (1) to come fairly close to capturing what we mean by 'well-being' in ordinary language and (2) to attain extensional adequacy (that is, to come up with a welfare theory that entails conclusions about what is or isn't good for someone that coincides, *in almost all cases*, with our pre-theoretical, considered judgments about what is or isn't good for someone). I think that the arguments that I have made here in Chapter 1 show that hypothetical DF theories cannot attain extensional adequacy.

<sup>92</sup> See Sumner 1996, 45-46.

<sup>93</sup> See Sumner 1996, 45, n. 1.

<sup>94</sup> See Brink 2006, 391.

<sup>95</sup> See Kagan 1998, 40.

<sup>96</sup> The idea here is *not* that someone's own human nature is some entirely non-normative/non-evaluative feature of herself – rather, the idea here is that someone's own human nature is normative/evaluative at least in the sense that it is receptive to those things that are pro tanto perfective of it (and, by extension, of herself taken as a whole).

<sup>97</sup> See Murphy 2001, 78.

<sup>98</sup> Finnis is clear that his OL theory is a perfectionist account (see Finnis 1998, 90-94). With respect to the goods that he puts on his list, Finnis says: "Everyone who understands the relevant sense of the word 'good' understands it in the sense of 'good for', 'beneficial for', 'more or less perfective of'" (see Finnis 1998, 91). Here it is Finnis's equating of 'good for' and 'more or less perfective of' that clearly signals that his OL theory is a perfectionist account. As for Kraut, he doesn't like the label 'OL theory', but he admits that he is an OL theorist (see Kraut 2007, 130, n. 50). Kraut prefers the label 'developmentalism' for his view (see Kraut 2007, 130, n. 50). Moreover, Kraut is a perfectionist OL theorist – on this point see what I say in Chapter 3, section 30. (One caveat here: Throughout I will be assuming that Kraut is an OL theorist, and this assumption is, I believe, correct. But it is worth noting that Kraut occasionally makes comments that suggest that he is a hybrid theorist who maintains that something can be an aspect of one's welfare only if she enjoys it – for instance, see Kraut 2007, 126-130. I don't think that this claim – the claim that

---

one must enjoy X in order for X to count as an aspect of one's welfare – fits well with the rest of Kraut's project. Also, Kraut himself doesn't say much, if anything, about this claim during the part of his book in which he is advancing the core of his own welfare theory – see Kraut 2007, 131-204. Again, then, I will be assuming throughout that Kraut is in fact an OL theorist and not a hybrid theorist of any sort.)

<sup>99</sup> For a discussion of accomplishment, see Griffin 1996, 19-21. By 'accomplishment' OL theorists aren't referring to trivial doings/actions – in effect, OL theorists use the word 'accomplishment' to mean '*worthwhile* accomplishment'.

<sup>100</sup> It might be wondered why this fine-grained approach is acceptable for OL theorists, but not for DF theorists. DF theorists may say: 'If we pay closer attention to what people really want, we will see that they never want things that, from the standpoint of common sense, we are prepared to deem bad for them. Rather, people only want the good things – that is, the things that, from the standpoint of common sense, we are prepared to deem good for them.' But it is hard to see how a DF theorist – or at least any typical DF theorist – could legitimately say this. After all, DF theorists typically conceive of our desires as being blind (or spontaneous) occurrences: DF theorists typically conceive of our desires as spontaneously fixing on value-free objects and thereby making these objects be prudentially valuable. To say that our desires zero in on the positive – and only the positive – aspects of what is in the world is to assume that our desires are anything but the blind (or spontaneous) occurrences that DF theorists typically assume them to be.

<sup>101</sup> For a discussion of this objection (and a reply to it), see Griffin 1996, 19-36, especially 30-31.

<sup>102</sup> See Griffin 1996, 31.

<sup>103</sup> See Darwall 1992, 157.

<sup>104</sup> See Feldman 2004, 19-20.

<sup>105</sup> See Korsgaard 2004, 90, n. 27.

<sup>106</sup> So as to avoid confusion, it helps to keep in mind that there might be plenty of cases on OL theories where (1) X's being intrinsically good for A crucially depends on A's having some sort of pro-attitude toward something and yet (2) X's being intrinsically good for A does not at all directly depend on A's having any sort of pro-attitude toward X.

<sup>107</sup> See Darwall 1992, 155.

<sup>108</sup> See Williams 1981, 101-113.

<sup>109</sup> See Sumner 1996, 42.

<sup>110</sup> See Sumner 1996, 165.

<sup>111</sup> See Hurka 1993, 17-18.

<sup>112</sup> See Wolterstorff 2008, 234.

<sup>113</sup> See Rosati 1996, 297-326 (especially 313-315).

<sup>114</sup> See Rosati 1996, 313-314.

<sup>115</sup> See Kraut 2007, 9.

<sup>116</sup> One might say to me: 'But don't you have to deny that plants have a welfare? After all, you've just said that prudential value is an especially personal sort of value. And plants are not persons.' However, what I

---

meant was that prudential value *for humans* is an especially personal sort of value. There might be a coherent concept of prudential value *for plants* – nothing that I have said rules this possibility out.

<sup>117</sup> In her article “Internalism and the Good for a Person” Rosati argues for the thesis of prudential internalism in various ways. One of her arguments focuses on our interpersonal justificatory practices concerning well-being. Some of what I have to say about this matter here (in section 20) at least loosely follows what Rosati says there (see Rosati 1996, 315-320).

<sup>118</sup> See Murphy 2001, 118-126.

<sup>119</sup> See Griffin 1996, 35.

<sup>120</sup> See Rosati 1996, 301.

<sup>121</sup> Go back to our interpersonal justificatory practices concerning well-being. In justifying to A that X is part of her welfare, we will try (at least implicitly) to bring out a direct connection between X and A’s desire-set. Moreover, it is A’s desires *as they are* that matter here. Our aim won’t be to bring out a direct connection between X and A’s desires *as they would be* if things were in some way different than they are now. Also, when it comes to intrapersonal prudential deliberation, we don’t typically worry about the desires that we would have if certain non-actual conditions were to obtain. Indeed, when an individual engages in intrapersonal prudential deliberation, the norm, as far her desires go, is that she simply considers the ways in which, and the extent(s) to which, the various prudential options that are open to her hook into her desire-set – that is, into the set of desires that she *actually* has. My point here, then, is this. If at all possible, we should avoid the move to a hypothetical version of prudential internalism. It doesn’t fit well with the way in which we typically think about, and talk about, human well-being.

<sup>122</sup> See Rosati 1996, 301.

<sup>123</sup> See Rosati 1996, 301.

<sup>124</sup> See Rosati 1996, 303.

<sup>125</sup> See Rosati 1996, 307.

<sup>126</sup> See Rosati 1996, 309.

<sup>127</sup> See Rosati 1996, 304-305.

<sup>128</sup> See Rosati 1996, 306-307, where she says: “The requirement that a person under ordinary optimal conditions *regard* removed counterfactual conditions as appropriate seems too strong, however, since it apparently requires that she be able to make certain theoretical judgments. It is the job of the theorist, not of the person who is contemplating her good, to make such judgments and to design the proper counterfactual conditions. Still, while we may have no particular antecedent views about what conditions are appropriate, we clearly do care about our counterfactual concerns under some conditions and not others. All else being equal, we do care about whether we would enjoy something if we were able to experience it, and we do not care about whether we would enjoy it if we were brainwashed to do so. In keeping with the internalist effort to link motivation and value, we might say this: counterfactual conditions C are appropriate only if the fact that a person would come to care about something X for her actual self when under C is itself something that she would care about when under ordinary optimal conditions. A person need not care about X itself while under ordinary optimal conditions in order for X to be good for her. But if her good is not to be alienated, the fact that she would care about X for her actual self under a particular set of counterfactual conditions had better be something that would prompt her concern under ordinary optimal conditions.”

<sup>129</sup> See Rosati 1996, 305.



---

<sup>130</sup> See Rosati 1996, 307-308.

<sup>131</sup> There is a question that arises here: If hypothetical DF theorists go with a form of two-tier internalism, then will this help them in solving the alienation problems (that is, the problems with the *for* part of ‘good for’) that I discussed in section 15? Going with a form of two-tier internalism might (to some extent) help hypothetical DF theories in this regard, though I’m not sure it would solve the problem. Also, it would bring on another problem. As I noted at the end of section 12, part of the rationale for moving to a full information condition is to ensure sufficient broadness: In order to ensure that an individual really does hypothetically desire everything that is prudentially fit for herself (or everything that truly is an aspect of her well-being), it is natural to think that she must be informed about everything that is out there. If hypothetical DF theorists go with a form of two-tier internalism, and if the agent in question (whoever she may be) refuses to endorse counterfactual conditions C that involve full information, then this is going to lead to sufficient broadness problems for hypothetical DF theories.

<sup>132</sup> See Finnis 1980, 59-99.

<sup>133</sup> See Griffin 1996, 19-36.

<sup>134</sup> See Gomez-Lobo 2002, 6-40.

<sup>135</sup> See Finnis 1980, 88-89. And see Gomez-Lobo 2002, 17-18.

<sup>136</sup> See Griffin 1996, 29-30; see Finnis 1980, 86-87; and see Gomez-Lobo 2002, 10-13.

<sup>137</sup> See Brink 2006, 390-391.

<sup>138</sup> Someone might say to me: ‘Ordinarily, a desire explains an action, but you are here talking about general desires (say, the general desire for aesthetic experience) that explain not *actions*, but rather more specific *desires* (say, the desire to see beautiful paintings). You must be working with a non-standard conception of desire – indeed, isn’t it best to say that you are really speaking not of general *desires* here, but rather of something like general *tendencies* that we have in our own respective natures.’ First, I should stress that, indirectly, the general desire for, say, aesthetic experience does explain *actions* that, for the agent in question, instantiate aesthetic experience – so it’s not as though I am allowing for any complete severing of the link between desire and action. Second, I don’t think that there is anything non-standard about this. We are perfectly familiar with general desires (say, the desire to make something of oneself) giving rise to more specific desires (say, the desire to make something of oneself as a businessman). So, no, speaking of general tendencies rather than of general desires does not seem to be called for.

<sup>139</sup> In posing an alternative to my view, someone might say: ‘Typically, things work as follows: Jerry doesn’t want friendship – he just wants friendship with Janelle. This seems to be the way that things usually work: People just find themselves with desires for instances of the basic goods – there are no general desires for the basic goods that are present and that enter into the full explanation for why the individual desires this or that instance of this or that basic good.’ Naturally, I disagree. Suppose that Jerry tells us that he wants friendship with Janelle. And now suppose that we say to Jerry: ‘And you don’t want friendship in general? All you want is friendship with Janelle?’ And suppose that Jerry answers: ‘Right, I don’t care about friendship – I just care about friendship with Janelle.’ To my mind, this sounds crazy. Part of the reason Jerry wants friendship with Janelle is that he wants friendship. I would say the following to Jerry: ‘Look, what you want is friendship *and* friendship with Janelle. Part of the reason that you formed the desire for friendship with Janelle is that you *already* had a desire for friendship. Part of the reason you are motivated to be friends with Janelle is that you have a more general motivation to engage in friendship. This isn’t to say that there is nothing about Janelle in particular that makes you want friendship with her – this is just to say that, if we want fully to explain why you desire to be friends with Janelle, then we should reference your having a general desire for friendship.’ Perhaps the easiest way to see that the desires for the basic goods are present within people is to think of cases where instances of basic goods are taken away from someone. Consider this. When I went to college at age eighteen, I left home and was all alone for the

---

first time in my life. None of my friends from high school went to my college. For the first time in my life, then, I was in a situation where I had no friends around. What did I do? I made an effort to become friends with people. During my freshman year I became close friends with three guys who lived in my hallway in my dorm. They are still my best friends from college – in fact they are the only friends from college that I still keep up with. Was there something about them in particular that caused me to desire to become friends with them? Yes, there almost certainly was. But still, the full explanation for why I became friends with them includes a reference to my having a general desire for friendship. Part of my motivation to become friends with *them* flowed from my motivation to have friends *period*. It seems to me that this is how things typically work. Granted, we might think that we just find ourselves with desires for instances of basic goods. But we need to dig deeper than that. That is too superficial a way of looking at the matter. The right thing to think is that particular desires for instances of the basic goods are ones that have flowed from more general desires (that is, from desires for the basic goods themselves). I am not saying that any of this has to happen consciously or deliberately. After all, it is not as though people usually sit there and consciously think to themselves: Well, I desire friendship in general, and so now I am going to go out and find some friends. These things rarely happen on a conscious, deliberate level. Indeed, the norm is that someone has a desire for, say, friendship that she isn't conscious of, and then she forms more particular desires to be friends with particular people, where these more particular desires are ones that have flowed out of her more general desire for friendship.

<sup>140</sup> The literature on evolution is full of references to desires that have been *programmed into us* by natural selection (for instance, the desire to pass on one's genes and the desire to care for one's offspring). And it could be that our desires for the basic goods have been *programmed into us* by natural selection.

<sup>141</sup> For instance, it may be that some sort of an Aristotelian metaphysics – one according to which humans by nature desire the basic goods – is true.

<sup>142</sup> In characterizing the standard theory of desire, Schroeder says that the standard theory straightforwardly allows us to desire things that are “abstract”, and here Schroeder brings up the example of the desire “to make something of oneself” (see Schroeder 2004, 12). This is surely a desire for a very general thing, just like the desire for friendship is a desire for a very general thing.

<sup>143</sup> See Gomez-Lobo 2002, 35-37.

<sup>144</sup> See P. Gilbert 2001, xix.

<sup>145</sup> See p. 176 in *The Encyclopedia of Phobias, Fears, and Anxieties*. Also, see P. Gilbert 2001, 6 – where Gilbert describes cortisol as a stress hormone.

<sup>146</sup> See pp. 176-177 in *The Encyclopedia of Phobias, Fears, and Anxieties*. Also, see P. Gilbert 2001 – in particular, p. 6.

<sup>147</sup> See P. Gilbert 2001, 6 and 14.

<sup>148</sup> See P. Gilbert 2001, 4.

<sup>149</sup> See Stocker 1979, 744.

<sup>150</sup> See Smith 2004, 120-121.

<sup>151</sup> See Hollon 2006, 133-134.

<sup>152</sup> See Solomon and Haaga 2004, 13.

<sup>153</sup> See Solomon and Haaga 2004, 13.

---

<sup>154</sup> See Solomon and Haaga 2004, 13.

<sup>155</sup> See Solomon and Haaga 2004, 23.

<sup>156</sup> See Hollon 2006, 147.

<sup>157</sup> I went through depression when I was twenty-one. When I went to therapy, my psychologist said roughly this. “First, we try what worked in the past. When that doesn’t work, we get superstitious. And, when that doesn’t work, we get help.” And, for me, that certainly was how things had gone. The superstitious activities of depressives can be very bizarre. Gilbert, for instance, notes that during his depression he would sometimes sit in lukewarm baths at four in the morning (see P. Gilbert 2001, xix).

<sup>158</sup> Also, many depressives refuse help when others offer it to them. Here consider the following quote from *The Encyclopedia of Phobias, Fears, and Anxieties*: “Family, friends, and coworkers of depressed people may become frustrated with victims of depression because their efforts to help and comfort are of no avail. The depressed person will not follow advice, refuses help, and denies comfort” (176).

<sup>159</sup> See Thurstin 1999, 13.

<sup>160</sup> See Wagner 1999, 107.

<sup>161</sup> See Wagner 1999, 107.

<sup>162</sup> Anorexics, for instance, want to have perfect bodies. As Kinoy, Holman, and Lemberg put it, “A desire to perfect one’s self through one’s body, and by extension in every other way, is also a strong characteristic [of anorexia] and can supersede the reality of body structure and function, resulting in a distorted body image” (see Kinoy, Holman, and Lemberg 1999, 2).

<sup>163</sup> See Wagner 1999, 107.

<sup>164</sup> See Peterson and Mitchell 1999, 130.

<sup>165</sup> See Peterson and Mitchell 1999, 132.

<sup>166</sup> See Peterson and Mitchell 1999, 128. Also, see Bowers and Andersen 2007, 20.

<sup>167</sup> What the anorexic wants is to have a perfect body. Here consider a quote from Thurstin: “Several characteristics, including high achievement orientation, perfectionism, and acquiescence, are common in women who develop anorexia. Individuals with anorexia often desire to be ‘the best’ in everything, including school, recreational activities, and personal appearance. Pressure to achieve the ideal body may lead to preoccupation with even the most minor ‘flaws’” (see Thurstin 1999, 14).

<sup>168</sup> See Kinoy, Holman, and Lemberg 1999, 2.

<sup>169</sup> Baumeister notes that masochists don’t pursue pain in general. They only pursue it in certain situations; and even then everything has to be under control/planned out (see Baumeister 1989, 13).

<sup>170</sup> See Baumeister 1989, 13.

<sup>171</sup> See Grandin 2006, 79-80.

<sup>172</sup> With respect to avoiding other people, Grandin writes that, during her years of puberty, the teasing from other kids at school “was very painful” and that just the threat of teasing made her so fearful that she was “afraid to walk across the [school] parking lot because [she] was afraid somebody would call [her] a name” (see Grandin 2006, 78). Williams writes about her social interactions with those with whom she was taking

---

classes so as to get a university degree in education (so as to be able to teach). With respect to her classmates, Williams says: “Why don’t I know to follow groups outside of class? I noticed that we all got along and at first I went off to lunch with groups from my class who made me feel included. Then they seemed to drift off without me...In theory I know people often say things that are meant subtly to invite me to come along but people I know outside of the university know they have to make this very overt for me or I just take it as part of the conversation...I realize there are times I don’t want to go off with groups from my class but I think part of this attitude is fear. Another part is defensiveness because I don’t understand the cues and get left out, so I say I don’t want to be with them. They do like me and when they see me outside (when I happen to drift past people in my class) they call me over. I understand that and then I join them but I am finding myself increasingly left out and I am sure they think this is my choice (which is only half-true). This makes it even harder for me” (see Williams 1994, 78).

<sup>173</sup> See Williams 1994, 78.

<sup>174</sup> See Williams 1994, 79.

<sup>175</sup> See Donnellan, Leary, and Robledo 2006, 226.

<sup>176</sup> At least to a limited extent, both Grandin and Williams engage in friendship with their own families and also with people that they come across as they work (see Grandin 2006 and see Williams 1994).

<sup>177</sup> See Schreibman 2005, 28-31.

<sup>178</sup> See Hobson 2007, 141.

<sup>179</sup> See Hobson 2007, 148.

<sup>180</sup> See Hare 1970, 7.

<sup>181</sup> See Hare 1970, 7.

<sup>182</sup> See Hare 1970, 2-4.

<sup>183</sup> See Hare 1970, 5.

<sup>184</sup> See Ochberg, Brantley, Hare, Houk, Ianni, James, O’Toole, and Saathoff 2004, 343.

<sup>185</sup> See Ochberg, Brantley, Hare, Houk, Ianni, James, O’Toole, and Saathoff 2004, 343-344.

<sup>186</sup> See Ochberg, Brantley, Hare, Houk, Ianni, James, O’Toole, and Saathoff 2004, 344.

<sup>187</sup> See Edens, Marcus, Lilienfeld, and Poythress Jr. 2006, 132.

<sup>188</sup> See Edens, Marcus, Lilienfeld, and Poythress Jr. 2006, 132.

<sup>189</sup> See Edens, Marcus, Lilienfeld, and Poythress Jr. 2006, 141.

<sup>190</sup> See Guay, Knight, Ruscio, and Hare 2007, 707.

<sup>191</sup> See Medline Plus Medical Encyclopedia ([www.nlm.nih.gov/medlineplus/ency/article/000921.htm](http://www.nlm.nih.gov/medlineplus/ency/article/000921.htm)). This is a National Institutes of Health website. Here, in particular, look under the heading ‘Causes’.

<sup>192</sup> See Ochberg, Brantley, Hare, Houk, Ianni, James, O’Toole, and Saathoff 2004, 342.

<sup>193</sup> See Martens 2002.

---

<sup>194</sup> See Martens 2002.

<sup>195</sup> See Martens 2002.

<sup>196</sup> See Martens 2002.

<sup>197</sup> See Martens and Palermo 2005, 301.

<sup>198</sup> See Kraut 2007, 105.

<sup>199</sup> See Schroeder, 2004, 98.

<sup>200</sup> See Schroeder 2004, 146.

<sup>201</sup> I should note, for the sake of being totally forthcoming, that Schroeder works with a different definition of desire than I do. However, I don't think that this difference between us matters in this particular context.

<sup>202</sup> See Kraut 2007, 136, n. 4.

<sup>203</sup> See Kraut 2007, 131.

<sup>204</sup> See Kraut 2007, 136-137.

<sup>205</sup> I am tempted to say that there are certain non-standard cases – say, the case of friendship-less autistics – where perfectionist value and prudential value clearly come apart. The idea here would be this: 'Friendship isn't part of the welfare of friendship-less autistics, but friendship is part of the perfection of friendship-less autistics (since all humans are perfected by friendship).' Honestly, though, I am not sure that friendship is part of the perfection of friendship-less autistics. Since friendship-less autistics have no appreciable receptivity for friendship, it isn't clear to me that they can be perfected by friendship. It may be best, all things considered, to say that friendship is a far-possible (rather than a near-possible) component of the perfection of friendship-less autistics. If this is correct – if friendship is a far-possible component of the perfection of friendship-less autistics – then perfectionist value and prudential value do not come apart in the case of friendship-less autistics. My general point here is this: Appealing to certain non-standard cases may not help me if I am trying to show that perfectionist value and prudential value sometimes come apart.

<sup>206</sup> See Wall 2007.

<sup>207</sup> See Rosati 2006b, 108-109.

<sup>208</sup> See Finnis 1980, 92-95.

<sup>209</sup> For example, Russell Hittinger is a natural law theorist who has attacked Finnis's claim that there is no objective hierarchy among the basic goods (see Hittinger 1987, especially 76-77).

<sup>210</sup> I don't have a worked out theory of practical rationality. But, if there are any cases of conflict between what prudence and perfectionism prescribe, then, with respect to these cases, I would be inclined to say: Even if perfectionist prescriptions usually trump prudential prescriptions, there are times when, even if prudential considerations do not trump perfectionist considerations, it nonetheless is not practically irrational to obey prudential prescriptions rather than perfectionist prescriptions. On the other hand, if there are no cases of conflict between what prudence and perfectionism prescribe (which is what would follow if prudential value and perfectionist value always inhere in the exact same states of affairs, and to the exact same extents), then we can say this. As far as normativity goes, it does not matter whether one gives priority to prudence or to perfectionism, for, either way, one will get the exact same prescriptions. Thus the idea here would be this: There is one coin of normativity that has two sides, the prudence side and the

---

perfectionist side; and it doesn't matter which side of this coin of normativity you strive to obey, for, whether you follow prudential normativity or perfectionist normativity, you will get the exact same result.

<sup>211</sup> See Adams 1999, 83.

<sup>212</sup> See Adams 1999, 93-94.

<sup>213</sup> See Adams 1999, 95.

<sup>214</sup> See Adams 1999, 96.

<sup>215</sup> See Adams 1999, 96.

<sup>216</sup> See Adams 1999, 101.

<sup>217</sup> See Sumner 1996, 133.

<sup>218</sup> See Darwall 2006, 583.

<sup>219</sup> See Parfit 1984, 494.

<sup>220</sup> Sometimes welfare theorists claim that X's obtaining cannot advance A's welfare unless A *experiences* X's obtaining – though here, of course, a great deal hangs on what the word 'experience' means. My own stance on this matter is that questions concerning the experience of well-being are less important than the following question: Is the putative aspect of well-being (whatever it may be) something that is specifically *for the agent*? If the putative aspect of well-being is a relational value – that is, if it is indeed specifically for the agent – then that is enough to make it a candidate for being a genuine aspect of the agent's well-being. On the other hand, if the putative aspect of well-being is not a relational value (or, in other words, is not a value that is specifically for the agent), then the putative aspect of well-being cannot be a genuine aspect of the agent's well-being.