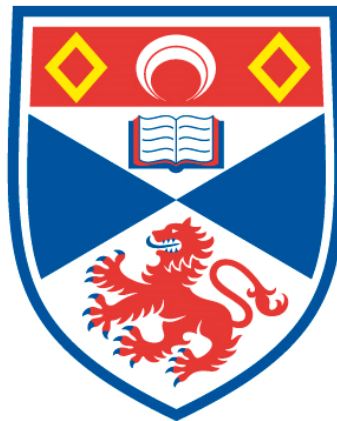


# THE ROLE OF WELL-BEING IN ETHICS

Raffaele Rodogno

A Thesis Submitted for the Degree of PhD  
at the  
University of St Andrews



2003

Full metadata for this item is available in  
St Andrews Research Repository  
at:  
<http://research-repository.st-andrews.ac.uk>

Please use this identifier to cite or link to this item:  
<http://hdl.handle.net/10023/13350>

This item is protected by original copyright

# **The Role of Well-Being in Ethics**

**Raffaele Rodogno**

**Department of Moral Philosophy**

**University of St Andrews**



ProQuest Number: 10166939

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



ProQuest 10166939

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

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

Th E470

(i) I, Raffaele Rodogno, hereby certify that this thesis, which is approximately 90000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

(ii) I was admitted as a candidate for the degree of Doctor of Philosophy in September 1999; the higher degree for which this is a record was carried out in the University of St Andrews between 1999 and 2003.

(iii) In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date

26/7/2003

Signature of candidate

(iv) I hereby certify that the candidate has fulfilled the conditions of the Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews.

Date

26/7/2003

Signature of supervisor

## Abstract

In this thesis I assess the role of well-being in ethics. In order to do so I reply to a threefold charge against the importance of well-being in ethics. In *What We Owe to Each Other* Scanlon argues (1) that the concept of well-being plays very little role in the thinking of an agent; (2) that no unified theory of well-being can be found; (3) that welfarism is false. In Part I, I argue that the concept of well-being does play an explanatorily and justificatorily important role in the thinking of a rational agent. I arrive at this conclusion by distinguishing levels of thinking activity as well as by considering the implicit rather than explicit role well-being plays in our deliberation. I conclude this part of the thesis by illustrating the relation between the idea of well-being, its parts and its sources. In Part II, I put forward a unified theory of *well-being* and I do so by taking on board with a slight modification Scanlon's own buck-passing account of value. I argue that something is a part of a person's good if, and only if, there is reason for this person to desire it. I claim that this account does not fall prey to the 'scope problem'. I also discuss a number of different though connected issues such as the defence of the claim that well-being is itself a normative notion and issues concerning the various parts of well-being. In Part III, I begin to sketch the normative role of well-being both first-personally and impartially. With Scanlon, I agree that welfarism is false. Yet, I argue in favour of a moderate form of welfarism, a view that takes a positive function of each individual's well-being to afford the ultimate criterion of practical reason.

## Preface

I acknowledge with immense gratitude John Skorupski and Sarah Broadie for their patient advice, support, and supervision. I would also like to express my gratitude to John Broome for his unfailing help. My manuscript has also been meticulously scrutinised, much to its advantage, by Brian McElwee. Discussion of various sections of the thesis with Paul Markwick has been of invaluable help. Special thanks go to Katrine Krause-Jensen in her role as a listener and valuable commentator. She and Davide Rodogno never failed to support me. I should like to record my gratitude to the School of Philosophical and Anthropological Studies, University of St Andrews, for its Doctoral Award (2000-2001) and to the *Fonds National Suisse de la Recherche* for its *Bourse jeune chercheur* (2001-2002). Most of all I would like to thank Anna and Daniele Rodogno for giving me the opportunity and indeed encouraging me to do that which I most desired to do.

# CONTENTS

<u>Introduction</u> .....	7
<u>PART I    <i>The Role of Well-Being in First-Personal Deliberation</i></u> .....	12
<u>Chapter 1    <i>The Alleged Unimportance of the Concept of Well-Being</i></u> .....	12
<u>1. Introduction</u> .....	12
<u>2. The importance of well-being the inclusive good</u> .....	12
<u>3. The concept of well-being and well-being</u> .....	18
<u>4. Scanlon's overall argument against well-being</u> .....	20
<u>5. Conclusion</u> .....	23
<u>Chapter 2    <i>Levels of Reflection and Categorially Different Goods</i></u> .....	25
<u>1. Introduction</u> .....	25
<u>2. Levels of activity and points of view</u> .....	25
<u>3. Intermezzo: some conceptual clarifications</u> .....	36
<u>4. Eudaimonia as an inclusive good</u> .....	39
<u>5. Conclusion</u> .....	45
<u>Chapter 3    <i>The Concept of Well-being and its Role in Practical Deliberation</i></u> .....	47
<u>1. Introduction</u> .....	47
<u>2. Mill, Scanlon, desiring and willing</u> .....	47
<u>3. Butler, self-love, the passions, and happiness</u> .....	56
<u>4. The idea, the parts, and the sources of well-being</u> .....	65
<u>5. Conclusion</u> .....	71
<u>Part II    <i>The Theory, the Parts, and the Limits of Well-Being</i></u> .....	74
<u>Chapter 4    <i>A Theory of Well-Being</i></u> .....	74
<u>1. Introduction</u> .....	74
<u>2. The epistemic role of desires</u> .....	74
<u>3. The semantic role of desires</u> .....	83



4. <u>Ideal-preference v. reason-to-desire</u> .....	89
5. <u>Conclusion</u> .....	97
<i>Chapter 5</i> <u><i>The Scope Problem and the Limits of Well-Being</i></u> .....	98
1. <u>Introduction</u> .....	98
2. <u>Darwall, the scope problem, and the normativity of well-being</u> .....	98
3. <u>The reason-to-desire theory and the scope problem</u> .....	111
4. <u>Well-being or personal good?</u> .....	115
5. <u>Posthumous well-being</u> .....	118
6. <u>Conclusion</u> .....	121
<i>Chapter 6</i> <u><i>The Parts of Well-Being: Some Remarks</i></u> .....	123
1. <u>Introduction</u> .....	123
2. <u>The list</u> .....	123
3. <u>Deep personal relations, virtue, knowledge, and achievement</u> .....	124
4. <u>Pleasure and happiness</u> .....	131
5. <u>Autonomy and its value</u> .....	136
6. <u>Conclusion</u> .....	141
<i>Chapter 7</i> <u><i>The Impossibility of Theories of Well-Being</i></u> .....	142
1. <u>Introduction</u> .....	142
2. <u>The charge against the possibility of theories of well-being</u> .....	142
3. <u>Scanlon's fixed points</u> .....	144
4. <u>Scanlon, choiceworthiness, and self-sacrifice</u> .....	154
5. <u>Conclusion</u> .....	162
<i>Part III</i> <u><i>Moderate Welfarism</i></u> .....	164
<i>Chapter 8</i> <u><i>Welfarism and Moderate Welfarism</i></u> .....	164
1. <u>Introduction</u> .....	164
2. <u>The buck-passing account: a first challenge to radical welfarism</u> .....	164
3. <u>From affective to practical reason: a second challenge to radical welfarism</u> .....	171
4. <u>Moderate welfarism: a partial defence</u> .....	172

<u>5. Conclusion</u> .....	179
<i>Chapter 9</i> <u><i>Moderate Welfarism, Morality, and Ethics</i></u> .....	180
<u>1. Introduction</u> .....	180
<u>2. From the good of one to the good of all: the impartial or general good</u> .....	181
<u>3. Morality the institution</u> .....	186
<u>4. Teleology, deontology and the epistemology of their reasons.</u> .....	191
<u>5. What things the impartial good cannot consist of</u> .....	195
<u>6. The corrective view, conservative holism, and moderate welfarism</u> .....	203
<u>7. Moderate welfarism: towards a complete defence</u> .....	209
<u>8. Conclusion</u> .....	220
<u>Conclusion</u> .....	222
<u>Bibliography</u> .....	224

## Introduction

Recently, Thomas Scanlon has claimed that well-being is commonly supposed to have the following features:<sup>1</sup>

1. Well-being is a *single* or *unitary* notion and is normally assigned the following three roles: (1) it serves as a basis for the decisions of a single rational being, at least as far as he alone is concerned (that is, leaving aside moral obligations and concern for others); (2) is what a concerned benefactor, such as a friend, has reason to promote; and (3) is the basis on which an individual interests are taken into account in moral argument. Of these three roles, the first is generally held to be of primary importance: well-being is important in the thinking of a benefactor and in moral theory because of its importance for the individual whose well-being it is. It is also commonly supposed that although the notion of well-being is important *for* morality, it is not itself a moral notion. It represents what an individual has reason to want for him- or herself, leaving aside concern for others and any moral restraints or obligations. Well-being is thus an input into moral thinking that is not already shaped by moral assumptions.
2. Well-being is also assumed to admit of at least rough quantitative comparisons of at least the following kinds: comparisons of the levels of well-being enjoyed by different individuals under various circumstances, comparisons of the increments in a single individual's well-being that would result from various changes, and perhaps also comparisons of the amounts of well-being represented by different lives, considered as a whole. It is taken to be an important task (important both for moral theory and for theories of "rationality" or "prudence") to come up with a theory of well-being: a systematic account of "what makes someone's life go better" that clarifies the boundary of this concept (the line between those things that contribute to a person's well-being and those that are desirable on other grounds)

---

<sup>1</sup> See Scanlon, 1998, Ch 3, esp. 108-112. The following points reproduce Scanlon's own words or give a very close paraphrase of them.

and perhaps provides a clearer basis for quantitative comparisons of the kind just mentioned.

3. It is commonly believed that value is teleological, i.e., all values are “to be promoted”. One way to defend this claim is to show that well-being is a “master value”: all other things are valuable insofar as they contribute to well-being.

Scanlon, however, is sceptical about the plausibility of these commonly held beliefs about the importance of well-being in ethics. Rather, he believes that:

1. Though individuals do have reason to be concerned with the things that contribute to their well-being, in regard to their own lives they have little need to use the concept of well-being. This concept has surprisingly little role to play in the thinking of the rational individual whose life is in question. Also, the notion of well-being used in moral theory is a notion shaped by moral theoretical concerns, not the intuitive notion that individuals might use in assessing their own lives.
2. We are not likely to find a general and unified theory of well-being on the basis of which one could see why the diverse things that contribute to well-being do in fact do so. Also, no account is likely to be found, capable (a) of delimiting the boundary of the concept of well-being—the line between contributions to one’s well-being and things one has reason to pursue for other reasons; and (b) of providing a standard for making more exact comparisons of well-being—for deciding when, on balance, a person’s well-being has been increased or decreased and by how much.
3. Well-being is not a “master value”.

The ultimate aim of this thesis is that of assessing the role of well-being in ethics. This is a rather wide topic, and though I do not take this thesis to exhaust all possible aspects of it, I believe that addressing Scanlon’s threefold charge against well-being will go a good way towards reaching the proposed aim. The thesis, then, is divided in three parts.

The aim of Part I is to assess the role of the concept of well-being in the thinking of a rational agent, and the alleged “moral” nature of the concept of well-being. In Chapter 1, I

present Scanlon's charge against the importance of the concept of well-being from a first person point of view, and the relevance of this charge to moral argument. Scanlon's argumentation heavily relies on two features: firstly, the fact that well-being is an inclusive good and secondly, the two criteria he uses in order to assess the importance of goods from a first personal perspective. In Chapter 2, I examine the Aristotelian distinction between the activity of statesmen and that of citizens, and the Aristotelian views concerning the formal nature of *eudaimonia*. Consideration of these topics will afford us material capable of reversing Scanlon's conclusions about the concept of well-being as a "moral" notion and begin to show the way in which the concept of well-being does play an important role in first-personal deliberation. This type of historically informed strategy also occupies part of Chapter 3. This time, however, I look at the philosophies of J. S. Mill and Joseph Butler. The ultimate aim of this chapter is that of providing an analysis of the concept of well-being. This analysis will provide us with an answer to this part of Scanlon's charge against well-being.

Part II is concerned with theories of well-being and the scope and limits of the concept of well-being. These two topics are intimately connected: if a theory allowed that rational pursuit of well-being can involve rational self-sacrifice then this theory can be considered not as a theory of well-being but as something wider, maybe as a theory of rational action. That's because this theory would single out a concept whose limits are too wide for it to be considered as the concept of well-being. In Chapter 4, I put forward what I will call the reason-to-desire theory of well-being: particular attention will be given to the role desires play in accounts of well-being. In Chapter 5, I determine what would be the scope and limits of well-being on our theory. It turns out that the concept of well-being is very wide in scope as a person's well-being would be contributed to by things as diverse as the well-being of other persons, all sorts of projects the outcomes of which (in one sense) do not involve the person, and even states of affairs that obtain after the death of the person whose well-being they contribute to. Yet, I will claim that this account of well-being does not include instances of rational self-sacrifice as contributions to a person's well-being. In Chapter 6, I discuss some of the issues concerning the various goods that our theory of well-being singles out as parts or constituents of a person's well-being. Finally, in Chapter 7, I shall apply Scanlon's charges against the possibility of theories of well-being to our own account as developed in the preceding three chapters.

In Part III, I assess the normative weight of well-being in ethics both under the artificial assumption that an agent's own well-being is the only well-being at stake and, more realistically, when an agent has to take into account the well-being of other individuals. In Chapter 8, I shall agree with Scanlon in rejecting the claim that well-being is a "master value". To say that, is to exclude what we shall call a radical form of welfarism. In the remaining part of the thesis, however, I will defend a more moderate form of welfarism which I will formulate as follows: there can never be most reason for an agent to do something that doesn't bring about the best outcome in terms of the impartial good. The bulk of Chapters 8 and 9 will consist in explicating and defending this thesis. Defending this thesis amounts to considering well-being as the ultimate standard of practical reason. In this sense, I will defend the view that well-being is of fundamental importance or plays a fundamental role in ethics.

In due course I will have much to say about the concept of well-being and its meaning. It is important however to make a short terminological note at this early stage. The notion of one's well-being is a general and unarticulated notion: it is the notion of one's interest, advantage, private or personal good, or benefit. This notion is articulated by the various things that contribute to well-being. To say that  $x$  is part of my well-being is to say that my obtaining of  $x$  would contribute to my well-being, be in my interest, etc. We could also say that  $x$  is good for me though this notion needs a further comment. As many have noted, 'good for' can refer to the specific functioning of organisms and artefacts: lubricant, for example, is good for the engine. Or again, my doctor may tell me that smoking is not good for me, meaning not good for the proper functioning of my organism. Yet, on the sense of 'good for' I intend to use, even if the doctor were right, smoking may still be good for me if, for example, it gave me a lot of pleasure. Finally, note that though something may contribute to my well-being, be in my interest or be good for me, it might not be the thing that maximised my well-being, most in my interest or best for me. When we talk of well-being in this latter sense, we talk of a person's good on the whole and we refer to a notion that extends to this person's whole life and, maybe, even beyond it.

The topic of this thesis is at the very core of ethical research. A thorough examination of the role of well-being in ethics is necessary to assess the relation between theories of the right and theories of the good (at least insofar as 'the good' is understood as a positive function of the well-being of all individuals), and the balance between egoism and

impartialism in ethics. Though in the last two chapters I *begin* to establish my own conclusions concerning these points, the bulk of this thesis is intended to be a preliminary to that project. This preliminary work, however, is also important in itself. Questions concerning the role of well-being in our first-personal thinking, the relation between the idea of one's well-being and the various goods that we often take to contribute to it, and the possibility of formulating a theory of well-being are very old questions, worth exploring whatever their usefulness for other purposes.

## **PART I    The Role of Well-Being in First-Personal Deliberation**

### **Chapter 1    The Alleged Unimportance of the Concept of Well-Being**

#### **1. Introduction**

Scanlon is particularly in disagreement with the idea that the concept of well-being is an important concept from the point of view of the agent whose well-being it is, as well as with the idea that well-being is a non-moral notion on which morality is grounded. In what follows, I examine these claims. I will focus in particular on Scanlon's way of testing the importance of a concept such as 'well-being' (§2), on the idea of importance itself (§3), and on the overall structure of Scanlon's argument in his chapter on well-being (§4).

#### **2. The importance of well-being the inclusive good**

According to Scanlon:

There are two related ways in which the importance of the concept of well-being in a given mode of thinking might be shown. First, it might be shown in the role that concept plays in explaining and helping us to understand the importance of the particular things that contribute to well-being. Second, it might be shown in the significance of the boundary of that concept—the difference it makes whether something is or is not a contribution to well-being.<sup>1</sup>

---

<sup>1</sup> Scanlon, 1998, 126.



Let us extract what we take to be Scanlon's two criteria for the importance of the concept of well-being:

(C1) The concept of well-being would be important if it played a role in explaining why in our thinking we take certain things to be important, namely, if we take them to be important because we take them to contribute to well-being.

(C2) The fact that the boundary of the concept of well-being is significant in someone's mode of thinking, i.e., that it makes a difference (in this agent's mode of thinking) whether something is or is not a contribution to one's well-being, is good evidence that well-being is important to an agent.

In the case of (C1), the concept of well-being must be the one we use in *explaining* why we take certain things to be important in our deliberation, i.e., we must make frequent appeals to the concept of well-being in order to explain why we take ourselves to have reason to desire and pursue many of the things we do. As for (C2), it is stipulated that the concept of well-being must have precise boundaries and that whether something falls inside or outside them has significant repercussions in our mode of thinking. That is, the fact that I conceive of  $\phi$ -ing as contributing or not contributing to my well-being makes a significant difference in my consideration as to whether I should or shouldn't  $\phi$ . Let us clarify further each of these two criteria.

### 2.1. The first criterion

Scanlon tells us that there are at least two levels of practical thinking at which the idea of one's well-being might be significant: either the level of everyday decisions about what to do or what particular goals to aim at, or the level of larger-scale decisions about how one's life is to go, such as what career to pursue or whether or not to be a parent. At each of these two levels, we may apply the two criteria in order to see whether the concept of well-being is or is not important. Here is an example of the first criterion applied to smaller-scale level decisions:

... it is certainly true that we have reason... to aim at things that contribute directly to our well-being, intuitively understood. We have reason to seek enjoyment, for example, to avoid illness and injury, and to do what will promote success in our aims. But the idea of well-being plays little if any role in explaining why we have reason to value these things. If you ask me why I have reason to listen to music, I may reply that I do so because I enjoy it. If you asked why that is a reason, the reply "A life that includes enjoyment is a better life" would not be *false*, but it would be rather strange. Enjoyments, success in one's main aims, and substantive goods such as friendship all contribute to well-being, but the idea of well-being plays little role *in explaining* why they are good. This might be put by saying that well-being is what is sometimes called an "inclusive good"—one that is made up of other things that are good in their own right, not made good by their contribution to well-being.<sup>1</sup>

This passage is of particular interest as it introduces the idea of well-being as an inclusive good. The example reads as follows: the concept of well-being is unlikely to feature in an agent's explanation of why he did a certain action even though it might be true that that action does, in fact, contribute to that agent's well-being. Therefore, according to (C1), there is no evidence that the concept of well-being is important to us. In order to make his claim intelligible Scanlon appeals to the notion of well-being as an *inclusive good*. That means that *well-being is not a separate good, but a good which is made up by other things which are good in themselves* or, at any rate, which are good not in virtue of their contribution to well-being. There are three remarks to be made about this passage from Scanlon.

The first remark regards the inclusiveness of well-being. Is Scanlon making a claim about (a) the nature of well-being, (b) the way we value well-being or (c) both (a) and (b)? I believe Scanlon is making a claim about both. More precisely, Scanlon explains (a) through (b). The fact that we value for themselves the various things that contribute to a person's well-being is evidence that well-being is an inclusive good. This is an important point insofar as in the rest of this thesis I shall not question the conclusion that well-being is inclusive in the way described by Scanlon.<sup>2</sup>

---

<sup>1</sup> Scanlon, 1998, 126-127, my emphasis.

<sup>2</sup> I will however qualify this claim by spelling out what it means to have reason to desire and pursue *for their own sake* the various things that contribute to a person's well-being. Also, in Chapter 2, I briefly consider the possibility of well-being as a dominant rather than an inclusive good.

Our second remark has to do with the stringency of Scanlon's claim that agents do not value the things that contribute to their well-being because they contribute to their well-being. Is Scanlon claiming that each thing that contributes to an agent's well-being is valued in itself and not because of its contribution to well-being? Scanlon is not as categorical as that. He writes that the idea of well-being plays *little* role—rather than *no* role—in explaining why we take things such as enjoyment to be good in themselves. Elsewhere, he writes:

On the one hand, when we say that something contributes to a person's well-being it sounds as if we are saying that it benefits him or her. But from an individual's own point of view many of the things that contribute to his or her well-being are valued for quite other reasons. From this point of view the idea of one's own well-being is transparent. When we focus on it, it largely disappears, leaving only the values that make it up.<sup>3</sup>

The transparency of the idea of well-being is explained by the fact that we *value* the goods that make up well-being for themselves. Scanlon, however, is not firm about the stringency of this transparency. In the last two sentences of this quotation he seems to be making two contradictory claims: the concept of well-being is said to be both transparent and *largely* transparent. Which of these two views, the more radical or the less radical should one attribute to Scanlon? When considering the role of well-being in deliberation over larger-scale choice Scanlon writes:

...the fact that it could make sense... to reject a life of devotion to some project because of the sacrifices in well-being that it would involve... seems to show that well-being is at least one important factor in such choices. Many of the things that contribute to one's well-being, such as health, enjoyment, and freedom from pain and distress, are certainly important factors in such a choice. The idea of overall well-being may also play a role, but this is less clear, in part because the notion of well-being that can be appealed to in this context is unavoidably abstract and indeterminate.<sup>4</sup>

Scanlon's thought is not unequivocal. I think, however, that the more radical claim that the concept of well-being is *always* transparent cannot be endorsed. Suppose that you offer me an unspecified reward on condition that I carry out a great labour (and that I have reason to

---

<sup>3</sup> Scanlon, 1998, 133.

believe you are not cheating me): do I have reason to carry out the labour? Most of us would think that the prospect of increased well-being, even when it comes in an unspecified guise, would provide anyone with a reason to carry out the labour. It seems undeniable that we prefer greater well-being to less. The fact that the notion of well-being appealed to here might be “abstract and indeterminate” does not make this claim any less true.<sup>5</sup>

My third remark has to do with the expression “things that are good in their own right” or in themselves or things that are desired and pursued for their own sake. Scanlon mentions enjoyment as such a good: “If you ask me why I have reason to listen to music, I may reply that I do so because I enjoy it.” A good deal of both Chapter 2 and 3 will explore the different meanings and ways in which something can be said to be good in its own right or desired and pursued for its own sake. I will be concerned with two issues in particular. The first has to do with the meaning of “for its own sake”. Though Scanlon is not explicit about this issue, on his view it seems that if we desire and pursue, say, enjoyment for its own sake, then we are not desiring and pursuing it for the sake of our well-being. As our discussion of Aristotle will show, however, we do not have to agree with that view. The second issue we need to clarify is the *way* in which enjoyment is a good in its own right. Consider this: is a beautiful work of art “good in its own right” as well? Surely, it is, though it seems to me we need an account of goodness capable of distinguishing aesthetic value from other kinds of value. If there are kinds of value, or ways in which things are good, what kind of “good in its own right” is enjoyment? It wouldn’t be farfetched to think that it belongs to the category of things that are *good for* someone. But if that is so, then it seems that the concept of well-being is still very much part of the picture.<sup>6</sup>

## 2.2. The second criterion

Let us now move on to the second criterion, (C2). Scanlon seems to be making at least two points. Firstly, that, in our practical thinking, we do not often need or use a notion of net overall well-being, i.e., “a notion that brings together and balances against one another all

---

<sup>4</sup> Scanlon, 1998, 131.

<sup>5</sup> It should be noted that Scanlon does not think that the notion of well-being is always or largely transparent from points of view other than the first-person one. See Scanlon, 1998, 138.

<sup>6</sup> I will develop this argument further in Chapter 3 §2. As made clear in the Introduction, I am assuming that ‘x is good for y’ is equivalent to ‘x contributes to y’s well-being’.

the disparate things that contribute to the quality of a life.”<sup>7</sup> What do we need to do if, for example, we must decide whether we should make some sacrifice in order to achieve a goal? According to Scanlon, we do not need to estimate the net effects on our well-being by determining the contribution that achieving this goal would make to our well-being, separating this from the other reasons for pursuing it, and balancing it against the cost in well-being that this pursuit would involve. Consider, for example:

the reasons which move us to promote the interests of our families and of groups or institutions with which we have other special relations. ... On the one hand, we would not want to think that we promote the interests of our friends, families, and institutions for “selfish” reasons, but, on the other hand, we would not be good friends or family members or loyal members of our institutions if we did not feel a loss to them as a loss to us. From a first person point of view, however, we have no reason to resolve this ambiguity by deciding *where the limits of our well-being should be drawn*. It is of course important to us—important in our moral self-assessment—that our concern for our friends and family is not grounded entirely in benefits they bring to us. But, given that we care greatly about our family or friends, *we have no need to determine the degree to which we benefit from benefiting them*.<sup>8</sup>

In our first-person practical thinking, we need not and do not determine the limits of our well-being. I may benefit from benefiting the people I care about, but it is not the fact that I stand to be benefited that motivates me to act. Thus, well-being can be a concept whose boundaries are not precise because they need not be, because even if they were, that would not make much of a difference to our mode of thinking. The second point Scanlon makes concerning (C2) follows from this conclusion and relies on a point reached while discussing (C1). Given that many of the things that contribute to well-being are valued primarily for other reasons (i.e., well-being is an inclusive good), it follows that the boundaries of the concept of well-being are blurred. But they would not need to be otherwise, as that would not make much of a difference in our practical thinking. The last example, Scanlon thinks, can be taken to provide an illustration of this second point as well. An agent may value benefiting her family or friends for reasons other than its contribution to her own well-being and yet it will be true that she would benefit by benefiting them. Or again, with Scanlon, we may believe that success in one’s rational aim

---

<sup>7</sup> Scanlon, 1998, 128.

will contribute to one's well-being and yet "we generally pursue these aims for reasons other than the contribution that this success will make to our well-being, and from a first-person point of view there is little reason to try and estimate this contribution."<sup>9</sup>

In conclusion, according to Scanlon, the concept of well-being is not important from a first-person point of view. That is so, firstly, because the concept of well-being does not help us understand or explain the importance of the particular things that contribute to our well-being. Secondly, because the boundary of the concept of well-being, i.e., whether a particular object or state of affairs contribute to one's well-being or not, is mostly insignificant to agents in their deliberation. What is more, given the particular nature of well-being as an inclusive good that boundary is bound to be blurred.

### 3. The concept of well-being and well-being

Scanlon's account may leave one wondering about what exactly is being assessed by him. Is he trying to understand the importance of *well-being* to agents or the importance of the *concept* of well-being to agents? The section where he discusses this topic is entitled "The Importance of Well-Being: First-Person Perspectives". One may assume, then, that Scanlon intends to assess the importance of well-being itself, rather than the importance of the *concept* of well-being. And yet, he concludes the section in the following manner:

I conclude, therefore, that the *concept* of one's overall well-being does not play as important a role as it is generally thought to do in the practical thinking of a rational individual.<sup>10</sup>

Since what may or may not be increased by our actions (or by the actions of others) is of course not the concept of well-being but rather a person's well-being itself, the thing which we are here to assess the importance of should be *well-being* rather than the *concept* of well-being. To demonstrate the unimportance of the *concept* of well-being to agents is not *eo ipso* to demonstrate anything about well-being itself.

---

<sup>8</sup> Scanlon, 1998, 129, my emphasis.

<sup>9</sup> Scanlon, 1998, 129.

Some confusion arises from the fact that Scanlon uses the notion of “importance” in what I take to be two distinct ways. Think back to Scanlon’s description of the first criterion. Scanlon writes that the “the importance[] of the concept of well-being in a given mode of thinking... might be shown in the role that concept plays in explaining and helping us to understand the importance[] of the particular things that contribute to well-being.” There are two occurrences of “importance”. To understand whether the two notions of “importance” are distinct we might ask whether they are both intended to have some normative content. It seems clear that Scanlon intends the second instance of “importance” to have some normative content. He thinks that we take something such as enjoyment to be important in this sense if we take ourselves to have reason to value it or to want it and pursue it.<sup>11</sup> As for the first instance of “importance”, Scanlon would say that the *concept* of well-being would be important\_ if we frequently appealed to it to *explain* why we take ourselves to have reason to want and pursue things such as enjoyment (i.e., things that are important to us in the normative sense). The importance\_ of the *concept* of well-being might have is of an *explanatory* kind, not of a normative kind, and according to Scanlon, the concept of well-being has no such explanatory importance from a first-person perspective.

The question now is: does Scanlon think that well-being itself (not its concept) is normatively important? Here is what Scanlon says:

The particular goods that make up well-being are important from the point of view of the individual whose well-being it is ... But the boundary between one’s own well-being and other aims is unclear, and we have no need to clarify it. ... We rightly view the world through a framework of reasons, largely shaped by the aims and values that we have adopted, and we rightly make particular decisions by determining what these reasons support on balance, not by comparing the net changes in our overall balance of well-being. Among these reasons are those provided by ideas of right and wrong, justice and moral values. These values constitute some of the individual’s most important “aims”... and ...

---

<sup>10</sup> Scanlon, 1998, 132.

<sup>11</sup> Henceforth we shall avoid using the expression “having reason to value something” and use instead “having reason to want (desire) and pursue something”.

they also play an important role in shaping a person's other goals, including the most comprehensive ones.<sup>12</sup>

The goods that make up one's well-being *are* normatively important to the agent from his perspective; agents have reason to aim at them. However, they are not aimed at because they contribute to well-being, as the concept of well-being is not explanatorily important. These goods are aimed at as ends in themselves. A concept that plays a large role in our practical thinking, in our "framework of reasons and values", will be able to have an influence on our actions and our life.<sup>13</sup> But that is not the case of the concept of well-being. On the other hand, Scanlon goes on to show, the concepts of right and wrong, of justice and moral values in general are normatively important ends of ours, and their concepts shape our lives in ways in which the concept of well-being does not.

#### 4. Scanlon's overall argument against well-being

Scanlon cautiously writes that the goods that make up a person's well-being are normatively important to the person whose well-being it is though not *qua* contributors to her well-being. That is shown by the fact that the *concept* of well-being is not explanatorily important from that perspective. What's the upshot of this argument? At the outset of Chapter 3 of *What We Owe To Each Other*, Scanlon spells out as follows the view that he intends to deny in that chapter:

[W]ell-being is important in the thinking of a benefactor and in moral argument because of its importance for the individual whose well-being it is. In particular, although the notion of well-being is important *for* morality, it is not itself a moral notion. It represents what an agent has reason to want for him- or herself, leaving aside concern for others and any moral restraints. Well-being is thus an input into moral thinking that is not already shaped by moral assumptions.<sup>14</sup>

---

<sup>12</sup> Scanlon, 1998, 136. Or again, in the concluding section of his chapter on well-being Scanlon writes: "From a first-person point of view, the things that contribute to (one's own) well-being are obviously important, but the concept of well-being plays little role in explaining why they are important, and the boundaries of this concept are not very significant." Scanlon, 1998, 142.

<sup>13</sup> For a more detailed exposition of this part of Scanlon's thought see Scanlon, 1998, Ch.1, section 10.

<sup>14</sup> Scanlon, 1998, 109.



Let us extract the following two claims from this passage:

- (a) Well-being is important in moral argument because it is important from a first-person point of view.
- (b) The concept of well-being is not a moral concept and it is not shaped by moral assumptions.

In the concluding section of the same chapter, Scanlon claims to have shown that:<sup>15</sup>

- (c) Even though the goods that make up well-being are important from a first-person point of view, the concept of well-being is not important from that perspective.
- (d) The concept of well-being is a notion shaped by moral theoretical concerns.<sup>16</sup>

It must be noted that claim (a) is about the “importance of well-being” both in moral argument and from a first-person perspective. Given Scanlon’s next claims (c) and (d) and in light of our preceding discussion, how are we to understand (a)? My suggestion would be this:

- (a') an individual’s well-being is normatively important in moral argument because the concept of well-being is normatively important from a first-person point of view.

I would take this thesis to mean something like the following: when acting morally, we have reason to promote/take account of (some positive function of?) people’s well-being because each person has reason to want her well-being. Through criteria (C1) and (C2), however, Scanlon would challenge (a'): her own well-being is not what each person typically takes herself to have reason to want and pursue from her perspective (though she

---

<sup>15</sup> Scanlon, 1998, 142-143.

<sup>16</sup> More precisely, Scanlon writes: “these conceptions of well-being will be moral conceptions, that is to say, they derive their significance, and to a certain extent their distinctive shape, from their role in the moral structure in which they figure.” Scanlon, 1998, 110. Scanlon also claims to have shown that well-being is not a “master value”, i.e., it is not the case that all things are valuable only insofar as they contribute to individual well-being. I shall postpone the discussion of this point until Chapter 8.

takes herself to want for themselves the various things that contribute to it). Therefore, it is not clear that (a positive function of) people's well-being is what we have reason to promote/take account of when we act morally. Scanlon, however, never goes as far as claiming that well-being *itself* is not normatively important from either the third-person or the moral perspective. He claims only to have shown that the *concept* of one's own well-being has a surprisingly small role to play in the thinking of the rational individual whose life is in question and (consequently?) that the *concept* of well-being that we use in moral argument must derive its shape from some moral theoretical assumptions.<sup>17</sup>

The conceptions of well-being used in philosophy –says Scanlon– are moral conceptions: they derive their significance, and to a certain extent their distinctive shape, from their role in the moral structure in which they figure. To illustrate his point Scanlon points to the conceptions of well-being used for the purposes of distributive justice in political philosophy. More specifically, Scanlon refers to John Rawls's primary goods and Amartya Sen's capability sets as examples of criteria of well-being that would not quite correspond to our intuitive first-person notion of well-being; they would not capture all of the dimensions along which we could assess how well someone's life goes. Two people enjoying the same primary goods, for example, may give completely different assessments of their respective lives and be justified in doing so. Scanlon explains this divergence by pointing to the fact that these criteria are supposed to measure only those aspects of a life that it is the responsibility of basic social institutions to provide for. On this point, one ought to agree with Scanlon: there is indeed a divergence between the aspects of a life that these criteria are intended to capture and a complete picture of how well a person's life goes. Contrary to Scanlon, however, I shall argue in Chapter 2 that there is no necessity to let one's moral theoretical assumptions and needs shape one's concept of individual well-being.

---

<sup>17</sup> Scanlon writes: "Well-being has its greatest significance from a third-person point of view, such as that of a benefactor, and, at least arguably, in our thinking about right and wrong. From both of these perspectives it remains true that... the things that contribute to a person's well-being are important because of their importance to the person. But the importance of well-being as a *category*, and the shape and importance of particular conceptions of well-being, derive from the distinctive concerns of a (certain kind of) benefactor, and from the special requirements of moral argument." Scanlon, 1998, 142. At this point it might be wondered why Scanlon lays so much stress on the issue concerning the moral or non-moral status of the concept of well-being. Scanlon is concerned with the charge of circularity that may be levelled against a moral theory such as his that is grounded on moral concepts. In his work, Scanlon intends, on the one hand, to show that even theorists who ground their moral theory on well-being are after all grounding it on a moral concept. On the other hand, Scanlon also tries to show how the circularity of his theory is not a vicious one but has to be understood as a positive form of holism. See Scanlon, 1998, Chapter 5.

If Scanlon is right, there is no such thing as a non-moral notion of well-being in which to ground moral argument. Our well-being is not something that we typically take ourselves to have reason to desire and pursue. The things that are normatively important from an agent's own point of view, are both things that contribute to his well-being as well as many other things, such as considerations of right and wrong and various forms of excellence. What makes these things worthwhile, however, is not the fact that they contribute to one's well-being.<sup>18</sup>

## 5. Conclusion

We began this chapter by spelling out Scanlon's two criteria for the importance of the concept of well-being from a first-person perspective, (C1) and (C2). Scanlon takes the concept of well-being to fail the test set by these two criteria. In short, it would seem that any inclusive good is bound to fail the test and well-being is such a good. In the remaining two chapters of Part I, I will show how the work of Aristotle, John Stuart Mill, and Joseph Butler can provide us with some elements to counter some of Scanlon's conclusions. More precisely, I will focus on two areas. Firstly, I will challenge the appropriateness of the conclusions Scanlon derives from the two criteria. More precisely, I will argue that care must be taken when applying the two criteria at the level of ordinary practical thinking.

Secondly, I will examine the inclusive nature of the concept of well-being. More precisely, I will examine the nature of the relation between the various things that contribute to a person's well-being and this person's well-being itself. Though Scanlon thinks that the concept of well-being is unimportant from a first-person perspective, he does extensively claim that certain goods contribute to a person's well-being. So there is an  $x$  such that

---

<sup>18</sup> According to Scanlon there are at least two classes of values that are worthwhile and are not grounded in well-being. Firstly, moral values: treating others fairly may make one's life as well as that of others go better, but that would not be the reason for taking this aim to be worthwhile. Rather, it is worthwhile because it is required by the more general value of treating others in ways that could be justified to them. It may be true that living in accordance to these requirements of right and wrong may make our life better but this does not give us reason to be concerned with this kind of value. The second example concerns the value of various forms of excellence. What makes these pursuits worthwhile is not their contribution to one's well-being but the fact that they constitute serious attempts to understand deep and important questions. See, Scanlon, 1998, 142-143.

various goods can be contributors to  $x$  and yet (a) we do not seem to make much use of the concept of  $x$  from a first-person point of view; and (b)  $x$  is not a separate good but an inclusive one. I would like to get a better grasp of the nature of this  $x$ . What kind of relation is there between this  $x$  and its contributors? Does the fact that we pursue its contributors for their own sake exclude that we, at the same time, pursue them as contributors of  $x$ ? Scanlon does not consider these questions. But as I intend to show in Chapters 2 and 3, answers to these questions have a direct bearing on Scanlon's conclusions concerning the importance of the concept of well-being. These answers will be found with the help of the three philosophers mentioned above. In the end, however, I will put forward my own conceptual analysis of well-being.

## Chapter 2      Levels of Reflection and Categorially Different Goods

### 1. Introduction

In what follows I help myself to a pool of concepts to be found in Aristotle's *Nicomachean Ethics* and in the most recent commentaries it generated. This chapter is not an exercise in Aristotelian exegesis. I do not question which of a number of views can be most plausibly attributed to Aristotle. Rather, I examine the philosophical cogency of some of the views rightly or wrongly attributed to Aristotle. In particular, I focus on two kinds of topics discussed by Aristotle and his commentators. The first has to do with the distinction drawn by Aristotle between the activity of statesmen and that of citizens. I will later try to bring this distinction to bear on Scanlon's criteria of importance (C1) and (C2) (§2).<sup>1</sup> The second topic has to do with Aristotle's views on the formal nature of *eudaimonia*. The overall aim here is to see whether it makes sense to hold, for some  $x$ , both that  $x$  is constituted of various goods pursued for their own sake, *and* that in aiming for those goods one aims for  $x$ . As we shall see, some Aristotelian commentators took *eudaimonia* to be such an  $x$ . Next, I consider whether well-being too can be a case of  $x$  (§4).

### 2. Levels of activity and points of view

Aristotle aimed his *Nicomachean Ethics* at statesmen. The book is intended to gain an understanding of the kind of life statesmen would want their citizens to live. The subject matter, then, is what life should ordinary individuals live, though the investigation is conducted from the point of view of would-be political leaders. Even though these

---

<sup>1</sup> (C1) The concept of well-being would be important if it played a role in explaining why in our thinking we take certain things to be important, namely, if we take them to be important because we take them to contribute to well-being. (C2) The fact that the boundary of the concept of well-being is significant in someone's mode of thinking, i.e., that it makes an important difference (in this agent's mode of thinking)

statesmen take the everyday experiences and thoughts of ordinary individuals as the starting point of their reflections about the highest good, they themselves are well-educated and reflective people. Through his two criteria, (C1) and (C2), Scanlon too seems to start from our everyday experiences and unarticulated thoughts about our lives. However, as we will show below, Scanlon wrongly assimilates the first-person point of view to the pre-reflective level and from there concludes that the concept of well-being is not important from a first-personal point of view. To that conclusion, however, he adds that some concept of well-being might be important in moral argument, but that that concept would not be the same as our intuitive first-person notion of well-being (think back to our discussion of Rawls's primary goods and Sen's set of capabilities in Chapter 1), as it would have inevitably been shaped by our moral theoretical concerns. In this section, I intend to show that the pre-reflective and reflective levels do not coincide exactly with respectively the first-personal perspective and what Scanlon calls the "moral perspective". We can reach reflective conclusions about what is normatively important first-personally that are not shaped by our moral theoretical assumptions.

### 2.1. Scanlon's "strange reply"

Remember Scanlon's example for the first criterion of importance:

If you ask me why I have reason to listen to music, I may reply that I do so because I enjoy it. If you asked why that is a reason, the reply "A life that includes enjoyment is a better life" would not be *false*, but it would be rather strange. Enjoyments, success in one's main aims, and substantive goods such as friendship all contribute to well-being, but the idea of well-being plays little role in explaining why they are good.<sup>2</sup>

We should focus on the "strange reply": "A life that includes enjoyment is a better life". As noted in Chapter 1, Scanlon is here in the process of showing the explanatory unimportance of the concept of well-being by showing how it would fail the test set by (C1). Typically, we do not explain why we have reason to desire and pursue our aims by

---

whether something is or is not a contribution to one's well-being, is good evidence that well-being is important to an agent.

<sup>2</sup> Scanlon, 1998, 126.

appealing to considerations of well-being. It would be “strange” to do so. Scanlon is right but for the wrong reason. The strange reply is strange because it is platitudinous. I do not express a platitude to someone else because I take it that they know it already. This point can be revealed by slightly changing the example. Suppose that you enjoy listening to music so much that it causes you to be tachycardiac. A friend who knows about your medical situation will ask: “Why do you listen to music?”. By that, however, he clearly means “Why do you listen to music given that it makes you miserable?”. At that point, you might be justified in saying “A life that includes the enjoyment of listening to music is a better life”, meaning that you prefer a life with some enjoyment but some suffering to a life with no enjoyment but no suffering. Your reply would not be strange.

Suppose now that the claim ‘enjoyment contributes to a person’s well-being’ were in fact true. Would the “strange reply” then be considered to be strange at any level of enquiry or for any context? Suppose that an agent is investigating what ends he has reason to aim at in his life, what things would make his life go well. Suppose, more in particular, that he is reflecting on the particular question as to whether he has reason to pursue enjoyment embodied in the activity of listening to music. After careful reflection, introspection, and debate with fellow thinkers, the agent concludes that “A life that includes enjoyment is a better life” and that he has reason to pursue it by listening to music whenever it does not conflict with his other rational aims. In this context that would not be a “strange reply” at all.

It might be remarked, however, that that would be a reply to a different question, a question along the lines of “is enjoyment something that would make your life go better?”. This takes us right to the core of our argument against Scanlon. The context in which Scanlon’s question is asked can be taken to be a first-person, short-term, everyday context. What about the context in which the strange reply no longer seems so strange? Is it a first-person context? Scanlon would have to say that it is not, otherwise he would undermine the conclusion he reached through (C1) that the concept of well-being is not important from that perspective. He would rather have to say that the concept of well-being that is being implicitly appealed to in a context in which the strange reply no longer seems strange is one shaped by moral theoretical assumptions. In the remaining part of this section, I will attempt to show that Scanlon would be wrong to think so. The answer would not necessarily be one shaped by moral assumptions; we are not asking, as for example Rawls

and Sen do, what goods should be justly distributed. We are simply asking what things would make our life go better. In order to prove this point, in what follows I will avail myself of some of the conclusions reached by Sarah Broadie in her discussion of Aristotle's *Nicomachean Ethics*.<sup>3</sup>

## 2.2. What level of activity?

According to Broadie, in the *Nicomachean Ethics* Aristotle:

... operates on two levels, using similar language on both. He begins by arguing that the supreme good is the statesman's objective, and goes on to investigate the nature of that objective. This second stage continues through the rest of the work, but in Book I it culminates in the definition of happiness as virtuous or excellent activity. Now, this activity is distinct from that of the statesman as such. The statesman aims, we are told, at making excellent citizens; but his ultimate aim, the measure of his success, is the excellent activity of individual citizens—which is to say their happiness. The statesman's own happiness does not appear at either stage, although we can infer that his fulfilment as statesman lies in achieving his goal as such. This is the happiness of citizens...<sup>4</sup>

There are then two levels or kinds of activity involved in the *Nicomachean Ethics*: that at which the statesman operates and that at which ordinary individuals or citizens operate. The *Nicomachean Ethics* is addressed to the statesman. His task is that of bringing about the conditions for his citizens' *eudaimonia*. In order for him to do that, it is necessary to understand what each citizen's *eudaimonia* consists in. The subject of the *Ethics* consists precisely in that. Broadie goes on:

Confusion of the levels is not surprising, since the difference between 'statesman' and 'ordinary individual' is more conceptual than real. Aristotle's ethical inquiry is meant to educate the statesman about his proper goal. So far as any of us partakes in such an inquiry with a view to making its conclusions tell in our own lives and the lives of those around us, we too are 'statesmen'.<sup>5</sup>

---

<sup>3</sup> See Broadie, 1991, 45-48.

<sup>4</sup> Broadie, 1990, 45.

<sup>5</sup> Broadie, 1991, 46.



One point we want to retain from this passage is that any individual could, at least in principle, engage in the ethical investigation into what constitutes his or her *eudaimonia*. This is very important. Yet Broadie believes that there is a *conceptual* difference between ‘statesman’ and ‘ordinary individual’. This, I think, is also true. What is not clear is what is entailed by this conceptual difference. Following Scanlon, I think we could look at this difference as a difference in the point of view or perspective occupied by the inquirer. It is not clear, however, what ‘points of views’ are involved and what the shift in points of view would involve. Let us examine this question further.

### 2.3. Scanlon, well-being, and points of view

Scanlon thinks that there are three points of view from which the concept of well-being might be taken to matter. That of the first-person, that of the benefactor and the moral point of view. In Chapter 1, we discussed the first-person point of view. Let us omit discussion of the benefactor point of view here and focus directly on the moral point of view. Scanlon thinks that a moral theory might rely on the concept of well-being in three ways:

First, the notion [of well-being] might figure in the content of moral requirements. For example, we may be morally required, at least in certain circumstances, to promote the well-being of others... Second, well-being might play a role in the justification of moral principles even when it does not figure in their content. A principle requiring us to respect a certain right... might be justified on the ground that its observance would promote individual well-being. ... Third, insofar as a moral theory needs to provide some justification for morality as a whole—some answer to the question “Why be moral?”—it might seem, again, that this is best supplied by showing how morality contributes to each person’s well-being.<sup>6</sup>

The point that Scanlon makes for each of these potential roles for the concept well-being in moral theory, is that the concept we would be dealing with in any of these moral theoretical roles, would not be the same as the intuitive notion of *individual* well-being. To put it simply: if *x* is our intuitive notion of first-person well-being, then the concept of

---

<sup>6</sup> Scanlon, 1998, 137.

well-being used in moral theory must be something other than  $x$ , call it  $y$ . What is more, as we will soon see,  $y$  would have to be a moral notion. The best illustration of this point can be found in the moral principles used by social justice theorists. The content of these principles involves overall assessments of how well-off various individuals are (they would need a concept of well-being fulfilling the first type of need in the quote above). Rawls's primary social goods and Sen's capability sets, for example, are narrower concepts than our intuitive ideal of well-being. On the other hand, the utility function underlying social choice theories would be a concept wider than our intuitive idea of well-being.<sup>7</sup> In short:

All these notions are shaped by moral ideas arising from the particular moral questions that they are supposed to answer: in the case of social choice theory by a conception of citizens' right to have their preferences taken into account in shaping social decisions, and in the case of Rawls's and Sen's accounts by ideas about the line between those aspects of individuals' situations that are the responsibility of social institutions and those that are properly left to individuals themselves.<sup>8</sup>

Scanlon rightly thinks that the notion of individual well-being is not co-extensive with the notions of well-being specific to social justice theories. But why would any one think that well-being should be the  $x$  that social justice requires to be fairly distributed? As Scanlon himself reminds us in a footnote, after all, Rawls and Sen did not intend to offer accounts of what makes a life better from the point of view of the person who lives it.<sup>9</sup> And yet, that is precisely what *we* are after. Scanlon, however, thinks that the argument he gave for social justice theories applies to moral theory more in general:

... what are employed in moral argument are generally not notions of well-being that individuals would use to evaluate their own lives but, rather, various moral conceptions to show how well-off a person is—that is to say, conceptions that are shaped by one or another idea of what we owe to and can claim from another. ... I believe that the conceptions of well-being that figure in moral thinking more generally can be expected to diverge in similar ways from the conceptions that individuals might use in assessing their own lives. Whether they diverge or not, however, these conceptions of well-being will be

---

<sup>7</sup> See Scanlon, 1998, 116-118 and 138-139

<sup>8</sup> Scanlon, 1998, p.139

<sup>9</sup> Scanlon, 1998, 385 n. 1. See Rawls, 1971, Section 15; Sen, 1992, Chapter 3.

moral conceptions, that is to say, they derive their significance, and to a certain extent their distinctive shape, from their role in the moral structures in which they figure.<sup>10</sup>

The claim that moral theories *in general* can only avail themselves of moral conceptions of well-being largely divergent from individual conceptions of well-being is not fully argued for by Scanlon. I think Scanlon is conflating two related questions: the question “what is the human good?” with questions such as “how do we fairly and or impartially weigh/take into account the good of each individual?” or again “what is that *x* that should be fairly distributed?” (primary social goods, according to Rawls, capabilities sets, according to Sen). Though treatment of the first question is normally necessary to answer at least some of the questions of the second type, these two types of questions are separate. This is most clearly seen in Aristotle. The statesmen have to (1) understand what is the good for human beings in order to (2) bring about the conditions for each individual to thrive. The second task might be one that may involve the kind of moral assumptions that seem to worry Scanlon. But these two tasks though related are separate.<sup>11</sup>

At this stage the remark concerning the more-conceptual-than-real difference between the statesmen level and the ordinary individual’s level can be fully appreciated. As an ordinary individual you may decide to reflect on what it is that makes your life go well, with an eye to making the conclusions of your reflection bear on your life. As an ordinary individual, you become a statesman for a while. Will your conclusions inevitably be shaped by moral assumptions such as concern for the just distribution of some basic goods? Would your conclusions be ones “which derive their significance, and to a certain extent their distinctive shape, from their role in the moral structures in which they figure” where “moral structures” are, as in Scanlon, structures of justice and relations with others? The answer is no, and that is because there still is an important difference between the statesman and yourself. You are here concerned with well-being understood as an *agent-relative* and universalisable value, *not* with well-being as an *agent-neutral* and universalisable value.<sup>12</sup> In other words, you are attempting to understand how each life is valuable to the person whose life it is, not how to act (first personally, or as an impartial administrator) given that each person’s life has impersonal value.

---

<sup>10</sup> Scanlon, 1998, 110.

<sup>11</sup> Some may regard the philosophical statesmen as engaged in *ethical* even if not *moral* thinking.

<sup>12</sup> A full explanation of this terminology will be broached in Chapter 9.

Finding an answer to the question “what is the human good?”, however, might require a level of reflection above that which ordinary individuals are used to employ in the daily conduct of their lives. Each reflective agent may begin to answer the question by examining his or her unarticulated experiences. The question, however, may become a broader and more abstract one. It would concern *human beings* in general, and their intersubjectively intelligible interests. This question would be one within philosophical anthropology. Contrary to what Scanlon might think, however, answering this question from a philosophic-anthropological point of view is not to make an appeal to any special set of moral assumptions.<sup>13</sup> From this point of view, we are simply trying to understand what things make a life fulfilling or good for the agent whose life it is. We are not yet concerned with what would be the fair way of considering each individual’s good, nor are we concerned with the just distribution of primary social goods.

#### 2.4. Scanlon’s contention

We need to consider exactly what is involved in Scanlon’s conclusion to the effect that the concept of well-being is not important from the point of view of the person whose well-being it is. If he means that we don’t usually make our moment to moment decisions by reference to our well-being, then he may well be right. If he means, however, that it makes no sense for an agent to reflect on the anthropological question so as to shape a better life for himself, then Aristotle would certainly disagree, and so would I.

True enough, Aristotle saw reflection as the task of the statesmen. I would think, however, that practical and normative questions such as “what things would make one’s life go well?” or “what things are good for one?” or even “is this end, activity, or action better for one than this other?” are recurrent questions for any individual with some capacity for reflection, and what is more these are typically questions asked with an eye to *praxis*. We see these questions often being asked in the newspapers: “are people in our kind of society

---

<sup>13</sup> The term “moral”, here, has to be understood in Scanlon’s sense. Some, however, may argue that this sense of moral is somewhat artificial – an artefact of twentieth century emphases in ethics or “moral” philosophy. If “moral” philosophy were mostly concerned with the Aristotelian question, we would find it natural to call assumptions concerning it, ‘moral assumptions’. Also, the individual who engages in thinking about the human good so as to improve his own life could be thought of as fulfilling the *duty* of self-improvement – a

too addicted to working and earning money?" As we shall see more in detail later in this chapter, the idea of well-being lurks behind each of these questions. Would it make sense to ask whether we are too addicted to work unless we suspected that too much work is not good for us? Everyone is faced with some choices that will partly determine one's short-term and long-term well-being. Should I go to the cinema tonight or should I keep working? How far should I pursue my career when it is not so compatible with other aims I have? What of my children, what should I want for them? What things should they have in order to have the best possible life? Starting from these basic questions, it is in principle open to each individual to engage in further reflection, though at its highest level, reflection on these issues is normally conducted by philosophers. What life is most fulfilling for a human being? Is it the pursuit of all and only rational aims? Or are there rational aims the pursuit of which will not make one better off? If so, can it ever be rational for me to pursue them?<sup>14</sup> In virtue of what is an action, an object, or a state of affairs better for one than any other?

## 2.5. Conclusion

Let us bring these conclusions to bear on our initial discussion. I started off by asking in what circumstances the "strange reply" would be considered strange. I put forward the example of the tachycardiac to show that even at an everyday level and from a first person perspective it might not be strange to give the "strange reply". Now, I think that examples of this kind are not rare; this undermines Scanlon's conclusion concerning the unimportance of the concept well-being by using his own criteria (C1) and (C2).

Scanlon, however, might still think that though undermined, the main gist of his conclusion remains untouched: the concept of well-being has surprisingly little role to play in the thinking of each agent. At this stage, we considered what exactly Scanlon takes the normative and practical force of this conclusion to be. It may be agreed with him that our

---

moral duty, many would say (e.g., Kant). In the rest of this section, however, I will use the term 'moral' the way Scanlon does.

<sup>14</sup> Rational egoists would deny it and so would metaphysical egoists. Rational egoism is the view that an action is rational just insofar as it would promote the agent's own interest (note the difference with ethical egoism: an action is morally obligatory just insofar as it would promote the agent's own interests). Metaphysical egoism claims that when the agent's own happiness or interest is correctly understood we will see that the good of others is, in the appropriate way, *part* of the agent's own good so that acting on other-

moment to moment decisions are not usually taken by reference to the concept of well-being. But one cannot infer from that that it makes no sense for an agent to reflect on the question of how to shape his life into a better life for himself. That question, we argued, is central and may come in different forms and at different levels of intellectual sophistication.

Though at its most highly reflective level this question is explored by philosophers, we argued, firstly, that this question need not be shaped by moral theoretical assumptions as Scanlon thinks. We are asking what human well-being is, not what would be the fairest positive impartial function of well-being or its fairest distribution. Secondly, we found that the difference between the ordinary individual's level and the statesman's level, i.e., the difference between a pre-reflective and a reflective level, is more conceptual than real. That is best understood when we realise that the pre-reflective level provides the ground or material from which given experience and reflection, the abstract ethical thinking of the reflective level can take off.<sup>15</sup> Thus, the pre-reflective level is incorporated into the reflective one.

It might be wondered at this point why anyone would bother with the reflective task. We may have a reasonable purely *theoretical* interest in the question of *understanding* what things it is reasonable for us to pursue. I think, however, that in most cases each individual engages in this type of enquiry with an eye to its practical applications. It ought to be noticed here again that the practical interest is not necessarily one having to do with moral issues of the type Scanlon seems to have in mind. An agent might be unsatisfied with the way his life is going and may want to find a way to improve it. But just as often, at least today, those who engage systematically on this type of reflection will want to gain an understanding of what things contribute to a person's quality of life, and then devise the social and political arrangements that could best promote the quality of life of every individual. We are not quite at the same practical level as Aristotle's statesmen but the task is indeed very similar.

---

regarding moral requirements is a way of promoting his own interests. For an interesting discussion of these definitions see Brink, 1992.

<sup>15</sup> Braodie, 1991, 47, makes a similar point.

Let us end this section with an important remark. To claim that the pre-reflective level provides the material for the reflective level is not to claim that whatever verdicts one arrives at the first level are also true at the next level. We have seen that with the “strange reply”. The fact that at this level “considerations of well-being” (as Scanlon would put it) do not explain many of our actions (if that is indeed true) does not establish that “considerations of well-being” are not important from an individual’s perspective. On reflection, this pre-reflective judgement may be revised. Let me give an example to establish this point.

Suppose you have a choice between going to the cinema or going to a jazz concert. Suppose, at a pre-reflective level, you take yourself to have a reason to pursue both these activities because you enjoy them. Someone like Scanlon might conclude from this as well as other similar examples that the concept of well-being is not important at this level. Suppose, however, that whenever faced with a choice between these two particular activities you always choose to go to jazz concerts, because you enjoy that more. It will then be remarked to you that you are appealing to a new principle, namely that (*ceteris paribus*) one has reason to do that which one enjoys more rather than less. But where in the concept of enjoyment can we find the idea that, *ceteris paribus*, more is better? Later on, I shall argue that this idea is not to be found in the concept of enjoyment alone but rather in the thought that enjoyment is desired under the idea of one’s well-being. Something similar to that is certainly involved in Aristotle’s concept of *eudaimonia*.<sup>16</sup> Any complete answer to this question can only be found at a level that can hardly be characterised as pre-reflective. Thus, *contra* Scanlon, even a good number of examples from the pre-reflective level cannot establish that the concept of well-being does not matter to the person whose well-being it is, as that judgement might have to be revised from a reflective standpoint. More to the point, the fact that from a pre-reflective point of view we pursue a number of intrinsic goods for their own sakes does not rule out any reflective conclusions as to their connection with well-being. This is the topic of the following sections.

---

<sup>16</sup> I am thinking about *eudaimonia*’s formal criteria, self-sufficiency and completeness. See Aristotle, *NE*, 1097 b 16-21.

### 3. Intermezzo: some conceptual clarifications

#### 3.1. *Eudaimonia* as a dominant or as an inclusive good

In this intermediary section, I will clarify a number of concepts that will be central to the following discussion. In Chapter 1, we concluded that an inclusive good, such as Scanlon takes well-being to be, has two relevant features: (1) it includes other intrinsic goods; (2) agents largely take themselves to have reason to pursue the included goods for their own sake and not because they contribute to their well-being.<sup>17</sup>

Some Aristotelian commentators take *eudaimonia* to be an inclusive good. At least on some readings, however, *eudaimonia* differs considerably from Scanlon's conception of well-being with respect to (1). It is not merely the case that *eudaimonia* includes other intrinsic goods. Rather, on these interpretations of Aristotle, *eudaimonia* would include *all* intrinsic goods: we could call *eudaimonia* an *all-inclusive* good. Scanlon does not conceive of well-being in these terms, as he thinks that there are intrinsic goods that do not contribute to well-being.<sup>18</sup> In §4, I shall examine some inclusivist accounts of *eudaimonia* and compare them with the second feature of Scanlon's inclusivist account of well-being.

I should mention here, however, that other Aristotelian commentators take *eudaimonia* to be a central or dominant good. On this view *eudaimonia* is a particular *substantive* good (e.g., the activity of *theoria*) that has a central or dominant role within a life.<sup>19</sup> The central good in question paradigmatically has *eudaimonia*: the centrality of this good in the context of a human life characterises that life as a happy (*eudaimon*) one. Some versions of the dominance view accept that the happy life, to *be* happy, requires other intrinsic goods

---

<sup>17</sup> The 'mostly' reflects Scanlon's hesitation on this issue. As illustrated in Chapter 1, he does think that, sometimes considerations of well-being, i.e., the fact that a certain course of action might be the one that most contributes to one's well-being might be a good reason for action.

<sup>18</sup> Note that some Aristotelian commentators would call what Scanlon calls an inclusive good, a *plural* or *composite* good. White S., 1990, Crisp, 1994b. Then they would claim that *eudaimonia* is both all-inclusive and plural. See Crisp, 1994b. That would amount to saying that there are no intrinsic goods outside *eudaimonia* and that there are several intrinsic goods included under the heading of *eudaimonia*. On another view, *eudaimonia* could be exclusive (the opposite of all-inclusive) and plural: there would be at least one intrinsic good outside it and several constituent goods within it. White, S., 1990, seems to hold that view.

<sup>19</sup> See Broadie's discussion, 1991, esp. 27-32.



beside the central good.<sup>20</sup> These goods stand in a particular relation to the central or dominant good: without *eudaimonia* these goods would lose much of their point.

On both interpretations, *eudaimonia* is not the same *kind* of good as respectively (a) the goods that it includes, and (b) the goods that are *not* central or dominant. *Eudaimonia* is a functionally or categorially different good. Scanlon did not envisage the possibility that well-being could be a dominant good and in order for us to argue against Scanlon on his own terms we shall not examine that possibility either. In defence of Scanlon, however, we shall say that to think that well-being could be a substantive master-end of the kind 'dominant' commentators take *eudaimonia* to be, would be too unlike what anyone nowadays means by well-being. Also, on the dominance view, the central good is identified with *one* particular end or activity (e.g., *theoria*). Is it plausible to think that each agent's well-being is identified with the same end or activity (whichever we decide it should be)? Suppose, however, that we do allow each rational agent to have his own or her own particular central activity. What arguments would there be to believe that each rational agent's well-being is structured around *one central* activity? It is simply not generally true that what makes a person's life go well is the pursuit of one *central* activity (though that doesn't exclude that some individuals' well-being might be identified with one central activity in that way). It is perfectly conceivable that an agent may reasonably pursue a number of things that are good for him and that are unconnected under any *one* particular end or activity. On the basis of these points, I would take it that well-being cannot be plausibly conceived of as a dominant good. Scanlon's failure to consider this possibility does not seem to have been a serious fault.

In §4, we will consider the interpretation of *eudaimonia* as an inclusive good. In fact, we will consider two interpretations of this view and assess whether they provide us with a model on which it makes sense to hold, for some *x*, both that *x* is constituted of various goods pursued for their own sake, *and* that in aiming for those goods one aims for *x*. Before broaching these issues, however, we should say a few words concerning the way in which the words '*eudaimonia*', 'happiness', and 'well-being' will be used in what follows.

---

<sup>20</sup> We will say more about this later on. Note that, as Ackrill remarks, there is at least the possibility of another kind of dominance view. He writes: "By 'a dominant end' might be meant a *monolithic* end, an end consisting of just one valued activity or good, or there might be meant that element in an end combining two or more independently valued goods which has a dominant or preponderating or paramount importance." Ackrill, 1980, 17. The dominance view we sketched in the text belongs to Ackrill's second type.

### 3.2. Eudaimonia, Well-being, and happiness: terminological issues

The standard translation of *eudaimonia* in English is ‘happiness’. In the sections below, whenever the discussion will be conducted on my own terms, I will refrain from translating *eudaimonia* in any way and will use the Greek term itself. The reason is simple: I doubt that Aristotle’s *eudaimonia* is equivalent to our term ‘happiness’ in either its everyday use or its philosophical use. In fact, as I will claim in Chapter 6, the use of the term happiness in ethics is liable to give rise to confusion. The observation that *eudaimonia* does not translate well as ‘happiness’ is not an original one. As early as 1884, in a footnote to his *Methods of Ethics*, Henry Sidgwick wrote:

Aristotle’s selection of *eudaimonia* to denote what he elsewhere calls “Human” or “Practicable” good, and the fact that, after all, we have no better rendering for *eudaimonia* than “Happiness” or “Felicity,” has caused no little misunderstanding of his system. Thus when Stewart ... says that “by many of the best of the ancient moralists... the whole of ethics was reduced to this question... What is most conducive on the whole to our happiness?” the remark, if not exactly false, is certain to mislead his readers; since by Stewart, as by most English writers, “Happiness” is definitely conceived as consisting of “Pleasures” or “Enjoyments.”<sup>21</sup>

Thus, according to Sidgwick, English writers, at least towards the end of the nineteenth century, ‘definitely conceived of [happiness] as consisting of “Pleasures” or “Enjoyments.”’ And, of course, that is not what Aristotle meant by the term *eudaimonia*, in either of its formal or its substantive guises. In *Methods of Ethics*, Sidgwick decides to resolve this ambiguity by referring to *eudaimonia* as “Well-being”. In fact, I think there is reason to object to even this move. ‘Well-being’ is not a better translation for *eudaimonia* than ‘happiness’. It might be thought that what Sidgwick had in mind was Aristotle’s formal definition of *eudaimonia* as the ultimate desirable end. If we want to allow for an understanding of ‘well-being’ such that it is possible to have reason to sacrifice one’s own well-being, however, it might not be wise to take ‘well-being’ as a direct translation of *eudaimonia*, as on Aristotle’s account *eudaimonia* is not something virtuous agents can rationally sacrifice.

## 4. *Eudaimonia* as an inclusive good

### 4.1. Ackrill and Irwin

John L. Ackrill's inclusivist interpretation of *eudaimonia* involves a relation of subordination between *eudaimonia* and the various activities that constitute it; yet, Ackrill claims, this relation is not necessarily a means/end relation. Ackrill is alluding to the constitutive or part-whole relation between *eudaimonia* and the goods and activities that are its parts, ingredients, or elements. These are Ackrill's own examples:

[O]ne may think of the relation of putting to playing golf or playing golf to having a good holiday. One does not putt *in order to* play golf as one buys a club in order to play golf; ... It will be "because" you wanted to play golf that you are putting, and "for the sake" of a good holiday that you are playing golf; but this is because putting and golfing are *constituents* of or *ingredients in* golfing and having a good holiday respectively, not because they are necessary preliminaries. ... That the primary ingredients of *eudaimonia* are for the sake of *eudaimonia* is not incompatible with their being ends in themselves; for *eudaimonia* is constituted by activities that are ends in themselves.<sup>22</sup>

One could choose the activity of putting *because* one wants to play golf; similarly, one could choose the activity of playing golf *for the sake of* a good holiday. The pursuit of these activities, however, is *not* instrumental to respectively playing golf or having a good holiday. Rather, it is *constitutive* of each of them. It seems then that the expression *for the sake of* does not necessarily indicate the presence of an instrumental relation. So far so good. Aristotle, however, thought that the parts of *eudaimonia* are chosen both for their own sake as well as for the sake of *eudaimonia*. If we take the example of the good holiday, we would have to say that playing golf is chosen for its own sake as well as for the sake of a good holiday. Now, playing golf is (for some) constitutive of having a good holiday. But what if it suddenly wasn't? I suppose we would have to say one should not

---

<sup>21</sup> Sidgwick, 1907, 92 n. 2.

<sup>22</sup> Ackrill, 1980, 19.

choose to play golf anymore. But, then, does it make sense to say that one should play golf for its own sake? Compare this with the case of something that is never chosen for the sake of something else, such as *eudaimonia*. Whatever else happens, we should always choose it. There is then a clear difference between things that are chosen only for their own sake, and things that are chosen for their own sake as well as for the sake of something else.

In effect, I am asking the following question: is the fact that *x* is chosen “as a constituent of happiness” compatible with the fact that *x* is chosen “for its own sake”? These two expressions would surely be compatible if they were close to being synonymous. Broadie indeed remarks that Ackrill uses the expressions ‘for the sake of happiness’ and ‘for its own sake’ almost as synonymous.<sup>23</sup> But, at least for Aristotle, this should not be the case, as he thinks that there is a difference between *eudaimonia* and the other intrinsic goods and the difference consisted precisely in that the former is always chosen for its own sake and never for the sake of something else while the latter are chosen for their own sake as well as for the sake of *eudaimonia*. Ackrill does not seem to have a good enough answer to my question.

Commentators on Aristotle wishing to explore the connection between choosing a good for the sake of happiness and choosing it for its own sake often turn to J. S. Mill.<sup>24</sup> According to Terence Irwin, Mill and Aristotle might agree insofar as they both claimed that “it is possible to desire some things for their sakes as parts of happiness”.<sup>25</sup> In more proper Aristotelian terms one should say:

... we can choose some things both for their own sakes and for the sake of happiness because we choose them as parts of happiness, and not just as instrumental to happiness...<sup>26</sup>

So far, Irwin does not seem to have gone beyond the point we reported Ackrill as making. To choose something for the sake of happiness does not mean that we are choosing it as an

---

<sup>23</sup> According to Broadie, Aristotle would not claim this, as he rather suggests that there are instances in which choosing some of the ingredients of happiness like pleasure and honour for their own sake may turn out to be incompatible with choosing them for the sake of happiness. See, Broadie, 1991, 55, n. 25.

<sup>24</sup> See Irwin, 1975, 341; Irwin, 1995, 290-2; White, N., 1995, 276-8.

<sup>25</sup> Irwin, 1995, 290. Aristotle and Mill would however be in disagreement if Mill defended the stronger thesis that “desiring X for its own sake is always the same as desiring X as a part of happiness.” Irwin, 1995, 290.

<sup>26</sup> Irwin, 1995, 291.

instrument to happiness; we might be choosing it as a part of happiness. Irwin seems to think that since we are not choosing the parts of happiness as instruments we are simply choosing them for their own sake. As shown above, however, that step might be unwarranted. There seems to be a difference between choosing something for its own sake *tout court* and choosing something for its own sake as well as for the sake of something else. Irwin, however, has more to say concerning this question. Consider the case of an intrinsic good such as friendship. What would be implied by the view stating that we choose friendship “for its own sake and for the sake of happiness *because* we choose it as a part of happiness”?

To take this view... is not to imply that we ought to be concerned for friendship, say, partly for its own sake and partly for its contribution to happiness. For the very features of it that make it choiceworthy for its own sake also make it a part of happiness; in choosing it for the features that make it a part of happiness we are also choosing it for its own sake. There are some features of friendship that would make it choiceworthy even if it did not belong to a life that achieved happiness; when we choose it for the sake of happiness, we choose it for the sake of those features.<sup>27</sup>

Irwin claims that the good-making features of friendship are identical with the happiness-contributing features of friendship. It seems to me, however, that someone like Scanlon might still ask: why on any token choice of friendship, is an agent choosing friendship? The agent may answer in at least three ways: (a) because friendship is worth choosing for its own sake (Scanlon would be satisfied with this answer); (b) because friendship contributes to happiness (substituting “well-being” for “happiness”, Scanlon would think we do not generally give this answer); (c) as Irwin would say, one chooses friendship for its own sake *because* he chooses it as part of his happiness. Look at (a) and (b), first. Clearly, the fact that my choosing friendship for its own sake always contributes to my happiness is not enough to show that on any occasion if I choose friendship under description (a), I thereby choose friendship under description (b) (Scanlon would still be able to press his point concerning the unimportance of the concept of well-being).

---

<sup>27</sup> Irwin, 1995, 291. Similarly, in the case of virtue Irwin says: “when we decide on something as a constituent of happiness, and regard it as contributing to happiness for this reason, we thereby decide on it for its own sake, choosing it as a good in itself. If this is right, then Aristotle’s account of virtue shows not only that he believes we can consistently choose something both for its own sake and as a part of happiness, but also that he believes we can show how something is worth choosing for its own sake by showing how it contributes to happiness.” Irwin, 1995, 291-2.

Thus, the only option Irwin has is description (c). According to Irwin, at the ontological level, there is one single feature  $F$  with two descriptions:  $F$  is the feature that makes friendship worthwhile;  $F$  is identical to the feature that makes friendship a part of happiness. At the motivational level (cf (c)), I choose friendship for its own sake (as a worthwhile good) *because* I choose it as part of my happiness. The ‘*because*’ indicates that at the motivational level what makes me choose friendship for its own sake is the fact that it contributes to my happiness. That this is indeed Irwin’s thought is confirmed by this further comment:

... when we decide on something as a constituent of happiness, and regard it as contributing to happiness for this reason, we thereby decide on it for its own sake, choosing it as a good in itself.<sup>28</sup>

Two remarks ought to be made at this point. Starting with our discussion of Ackrill, the expression “for its own sake” seems to have caused us some trouble. An ambiguity is at the source of this trouble. On a stronger understanding, ‘for its own sake’ can be understood to mean *irrespective of context*. On this understanding of the expression, something cannot be both chosen for its own sake and for the sake of something else. On a weaker reading, however, for its own sake means *not as a means to something else*. Irwin clearly cannot take on board the stronger meaning. When our reason to choose friendship for its own sake is that it is a constituent of happiness (as he says in the last quotation), then take away the idea that friendship is a constituent of happiness and you no longer have a reason to choose it for its own sake. This of course entails that neither friendship, nor any other constituent of happiness, is an intrinsic good irrespective of context. An example of a good chosen irrespective of context would be *eudaimonia* itself, something we always choose for its own sake and never choose for the sake of something else. We now have it, then, that choosing friendship for its own sake as a constituent of happiness, is simply not to choose it as a means to something else rather than choosing it irrespective of context.

Notice, however, and this is our second remark, that this reading of (c) is very far from what Scanlon would think, as he claims that we take ourselves to have reason to choose

---

<sup>28</sup> Irwin, 1995, 291.

various things (that contribute to well-being) for their own sake irrespective of their contribution to well-being. Let us develop this point further in the next subsection.

#### 4.2. Back to Scanlon: lessons from Aristotelian inclusivism

As we saw in §3, Scanlon thinks: (1) that well-being includes other intrinsic goods; and (2) that agents largely take themselves to have reason to pursue the included goods for their own sake and not because they contribute to their well-being. The crucial point here is (2). Contrary to Scanlon, Irwin thinks that when we choose something as a constituent of happiness (*eudaimonia*), and regard it as contributing to happiness for this reason, we thereby choose it for its own sake, not as a means to something else.

Let's translate Irwin's thought in the terms relevant to our discussion of Scanlon. According to Irwin, if I choose enjoyment as part of my well-being, and regard it as contributing to my well-being for this reason, I thereby choose it for its own sake. The deliberative route taken by Irwin's agent seems to run in exactly the opposite direction to that taken by Scanlon's agent. On Irwin's account, an agent considers as the reason for choosing enjoyment the fact that it is a part of his happiness; when he chooses enjoyment for that reason he chooses it for its own sake (in the weak sense of the expression). On Scanlon's account, the agent considers the fact that something is enjoyable to be itself the reason for choosing it; enjoyment is good in its own right. That enjoyment also contributes to one's well-being seems to be a fact largely irrelevant to the agent's deliberation; that fact is something of an epiphenomenon, motivationally and justificatorily largely inert.

This is a crucial difference and it is time to take a stance on this issue. With Scanlon, I agree that often when deliberating we do not seem to consider the fact that something is or is not a constitutive of our well-being. More precisely, I think that often, this fact is not considered *explicitly*. This, however, can be explained by the fact that most of our deliberation relies on direct and indirect experience and knowledge of what objects and activities are good for us, or part of our well-being. In our deliberation, then, more often than not we need not attend to this issue, as the fact that something is or is not good for us is part of the background knowledge in our deliberative process. Consider Scanlon's cinema example: "why do you take yourself to have reason to go the cinema? Because I enjoy it." Our well-being does not explicitly appear anywhere in our deliberation. Suppose,

however, that a friend of mine with insider knowledge tells me that in the last month the cinema I want to go to has received several bomb threats. In other words, my friend is telling me that going to the cinema might not be *good for me*. Sensibly, I take myself to have reason on the whole not to want to go the cinema. What would justify my choice not to go? Most of us would find it perfectly satisfactory to say that it is my well-being, i.e., the idea that it would not be good for me to go.<sup>29</sup>

Thus, though the idea that something is or is not part of my well-being is not always explicit in our deliberation, it definitively is an *implicit* part of it. As soon as, on the basis of our experience and information, we become unsure as to the outcomes of our choices it seems in most cases appropriate to ask the question “Is this good for me?” or “Is this the best alternative for me?”. In these cases, considerations of well-being become apparent and explicit.<sup>30</sup> My view on the role of well-being in our deliberation, then, is different from Scanlon’s though it is not the same as Irwin’s either.<sup>31</sup> Agents do not often *start* their deliberative process by considering whether what they choose is a part of their well-being. These considerations become explicit only in certain cases.

Finally, we ought to consider another one of the lessons our discussion of Aristotle brought to the fore. In what sense are we to use the expression “for its own sake” when we say that an agent has reason to want and pursue, say, enjoyment for its own sake? Scanlon does not

---

<sup>29</sup> This example might need further spelling out. Suppose, for example, that my friend is the bomber and knew that a bomb capable of disintegrating the whole block would go off at the time I intended to be at the cinema, just a few minutes before the session started. In this case there is no reason for me to desire to go to the cinema as nothing good for me would be obtained by going to the cinema. Next, consider the case in which the bomb would in fact go off just as the session reaches its end. In that case there would be some reason to desire to go to the cinema, as I would get to enjoy the movie, though there would be stronger reason to desire not to go, as that would not be overall good for me. This is the case I had in mind in the text above. It might be thought that the “bomb-threat” example is too extreme, though of course that would depend on what part of the world one lives in. The same point can be easily made with everyday type of cases. A child asks his mother: “Why do I have to go to school? I hate it!” The mother answers or thinks to herself “Because that is good/best for you”. Or again, my friend is trying to convince me to try some hash/a new machine at the amusement park/to take a course in culinary arts. It would be appropriate and natural on my behalf to consider how and whether each of these things would be in my interest or good for me.

<sup>30</sup> I am not here making the claim that it is always irrational to choose something we believe not to be good for us. I am not excluding that we may rightly consider ourselves to have most reason to do what is morally obligatory though it clashed with what would most be in our interest. Some, however, may want to say that deliberation about moral actions should not at all involve consideration of one’s own well-being. I believe this position is too moralistic. If there were two courses of actions to do what is morally obligatory and one demanded a much lesser sacrifice in my well-being shouldn’t I let this consideration count in favour of it? Suppose that in fact one course of action required no sacrifice in my well-being at all. Why should I not consider it and prefer it?

<sup>31</sup> Insofar as we can take Irwin’s conclusions about *eudaimonia* to bear on our discussion about well-being.



mention the possibility of there being a stronger and a weaker sense in which something can be pursued or chosen for its own sake. Surely, however, Scanlon cannot think that an agent has reason to want and pursue enjoyment for its own sake *irrespective of context* at least if Scanlon also thinks that enjoyment is a part of well-being. Think back to our cinema example above. Would I still be justified in taking myself to have reason to want to go to the cinema if I knew it wasn't good for me? As far as my view goes, I shall rest content with the weaker sense. Though there is reason to desire enjoyment for its own sake, there is no reason to desire any particular enjoyment unless the agent rightly regarded it as *good for* him or her or in his or her interest. This can consistently be asserted when "for its own sake" means "not as a means to something else" rather than "irrespective of context".

## 5. Conclusion

At the end of the last chapter, I listed a number of unanswered questions arising from Scanlon's discussion of the concept of well-being. The first set of questions concerned the two criteria Scanlon used in order to assess the importance of the concept of well-being in the thinking of an agent. In the first part of this chapter I concluded that what answers these criteria yield partly depends on the level of reflection at which they are applied and contended that Scanlon did not apply them at the right level. Contrary to Scanlon, I also showed that there is a first-person, non-moral notion of well-being that agents appeal to in their everyday life though, at a higher level of reflection, this notion is also used for moral theoretical purposes.

The other set of questions raised at the end of the last chapter had to do with the relation between the concept of well-being and the various things that contribute to well-being. I asked what was the nature of this  $x$  that is contributed to by the various things we have reason to desire and pursue in their own right and what kind of relation ties this  $x$  to its contributors. Does the fact that we pursue its contributors for their own sake exclude that we, at the same time, pursue well-being for its own sake? From our analysis of the inclusivist reading of *eudaimonia*, we learned that the expression "for its own sake" is liable to a weaker and a stronger reading. On the former reading, the fact that one takes

oneself to have reason to pursue a certain good for its own sake does not exclude that one could take oneself to have reason to pursue that good for the sake of his well-being.

Moving away from Aristotelian commentators, we discovered that the concept of well-being might implicitly be part of our deliberative process in a way unexamined by Scanlon. Even when we take ourselves to have reason to desire and pursue, say, enjoyment for its sake, we would take ourselves to have stronger reason not to do so if further information showed us that it might not be good for us to desire and pursue enjoyment on that occasion. In its weaker sense, we may still be able to say that we desire and pursue enjoyment for its own sake, under the idea of it as good for one. The upshot is that *contra* Scanlon the concept of well-being does play a significant role in the thinking of a rational individual. The relation between well-being and its constituents, however, needs to be examined in greater detail. That is the task of the next chapter.

## Chapter 3      The Concept of Well-being and its Role in Practical Deliberation

### 1. Introduction

The ultimate aim of this chapter is that of providing a final reply to Scanlon's claim concerning the unimportance of the concept of well-being from a first personal perspective. This reply will be expressly stated only in the concluding section (§5): it is not the case that the concept of well-being "has surprisingly little role to play in the thinking of the rational individual whose life is in question". More positively, my conclusion will be to the effect that for each and every final good constituting a person's well-being, remove the idea that this good contributes to this person's well-being and you will remove the idea of that thing as what this person has most reason to want. I will reach this conclusion by turning to the work of Joseph Butler (§3) as a source of inspiration for my own analysis of the concept of well-being (§4). First, however, we need to clarify a difference in the way Scanlon and I would analyse the various things that contribute to a person's well-being (§2).<sup>1</sup>

### 2. Mill, Scanlon, desiring and willing

#### 2.1. The buck-passing account of goodness

Scanlon considers enjoyment to be something we have reason to desire and pursue for its own sake. He would equally say that enjoyment is a good in its own right. That is because Scanlon endorses what he calls the buck-passing view of goodness. In what follows I shall myself endorse this account, though, as we are soon to see, I will endorse it with an important qualification.<sup>2</sup> On this view something is good if, and only if, it displays other

---

<sup>1</sup> Though the relevance of this difference is going to be most crucial to Part II and III it is convenient for expository reasons to introduce it at this stage of the thesis.

<sup>2</sup> In Chapter 8 I will defend the buck-passing account against an important charge.

properties that provide reason to have a certain pro-attitude towards it. “Exactly what these reasons are, and what actions and attitudes they support will be different in different cases.”<sup>3</sup> If, for example, the putative good in question is a work of art we may say that it is good if and only if it displays other properties that provide an agent with a reason to *admire* it. The attitude of admiration will in turn support certain kinds of actions such as those that *express* admiration.

Now consider the case of enjoyment. Agents rightly take it that objects or pursuits that display the property of being enjoyable are good. The question we need to examine, however, is whether these things are good in the same *way* in which a work of art is good. They clearly aren’t. These objects or pursuits do not primarily have properties providing reason to feel admiration. In other words, these objects or pursuits are not aesthetic goods. In what way are they good then? I can find no answer other than that they are prudentially good or *good for the agents* who attain these goods. This point may sound like a point against Scanlon: it seems now as if we have reintroduced the idea of well-being through the back door. On the contrary, however, the subtlety of Scanlon’s argument can be fully appreciated only now. Scanlon argues that just as we have reason to value in themselves the various things that contribute to our well-being we have reason to value in themselves things that do not contribute to our well-being. Hence, Scanlon would say, it is not important from an agent’s point of view whether something is or isn’t good for him.

In order for Scanlon’s argument to go through, however, there must be no difference in the way we have reason to value things belonging to these two different categories. In other words, the pro-attitudes the things that contribute to a person’s well-being give us reason to have must be the same as the pro-attitudes the things that do *not* contribute to a person’s well-being give us reason to have. Unsurprisingly, Scanlon seem to think that just as we have reason to *want* and *pursue* things that are good for us we have reason to *want* and *pursue* things that are not good for us. He also refers to both types of things as to what we have reason to *care about*.

Scanlon’s argument may be countered by individuating one kind of pro-attitude that all and only prudentially good things would give us reason to have. In what follows, I will claim

---

<sup>3</sup> Scanlon, 1998, 97.

that there is one such pro-attitude, namely our desires: something is good for an agent if and only if there is reason for an agent to *desire* it. I will take this account to be an instance of the buck-passing account of goodness. If am successful in this task, I will have shown that the goods contributing to an agent's well-being can be distinguished from the goods that don't, and that from the agent's own perspective. As we shall see, this point can be endorsed while maintaining that, by and large, an agent has reason to want the things that are good for him for themselves and under the idea of his well-being. The work of J. S. Mill will be central to develop this argument.

## 2.2. Mill: desiring and willing

In *Utilitarianism* Mill says:

There is in reality nothing desired except happiness.<sup>4</sup>

By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.<sup>5</sup>

Concerning the notion of a desire Mill says:

...desiring a thing and finding it pleasant, aversion to it and thinking of it as painful, are phenomena entirely inseparable, or rather two parts of the same phenomenon; in strictness of language, two different modes of naming the same psychological fact: that to think of an object as desirable (unless for the sake of its consequences), and to think of it as pleasant, are one and the same thing; and to desire anything, except in proportion as the idea of it is pleasant, is a physical and metaphysical impossibility.<sup>6</sup>

Mill seems to suggest that desiring something and finding that thing pleasant are two different ways of referring to the same psychological fact. It has been noted that there is some ambiguity in the expression "finding a thing pleasant".<sup>7</sup> It can mean feeling pleasure at the thought of doing that thing, or thinking of that thing as a pleasant thing to do, or

---

<sup>4</sup> Mill, 1998, 84, IV8.

<sup>5</sup> Mill, 1998, 55, II2. Firstly, note that this formulation is mute as to whose happiness is being desired. In *Utilitarianism* Mill seems indeed to be shifting from the agent's own happiness to general happiness. Secondly, note that Mill did not maintain *Motivational Hedonism*, i.e., the thesis that agents are always and exclusively motivated to act by a concern with their own happiness/pleasure.

<sup>6</sup> Mill, 1998, 85, IV10.

both. Mill must mean that whenever a person wants something he wants it in virtue of thinking it will be pleasant. If that was not so then Mill could not claim that nothing is desired except happiness, i.e., pleasure and the absence of pain. What is important here, however, is that according to Mill, to desire something is to conceive of something as pleasant. For our purposes, I would like to make two amendments to Mill's view. Firstly, it does not seem true that all things we desire, we desire under the idea of them as pleasant. Suppose I am sitting at the hospital waiting to hear whether I tested positive for AIDS. I clearly desire to know the result of my blood test though it is hardly plausible to think that I do so under the idea of pleasure. In other words, there seem to be other desirability features beside pleasure. In the example I just mentioned, the desirability feature in question might be taken to be 'knowledge of my situation'.

Next, focus on the last three lines of the quote. We can read the following two thoughts: (a) judging an object to be desirable for itself is the same as judging it to be pleasant; and (b) to desire something for itself is the same as judging it to be pleasant. We have already noted that we may desire things other than pleasure. Here, however, we should focus on (a) and (b) jointly, and note how they suggest that when we desire something we think of that thing as desirable. Once again, I here feel Mill's position to have been too radical. Is it really impossible to have a desire for something, knowing that that thing is after all not desirable? And couldn't someone find something desirable without thereby desiring it? I would however rest content with a weaker version of this Millian thesis: having a desire for something involves the tendency to see that thing as desirable, worthy of desire, or good to have under the idea of it as being pleasant or under some other desirability feature.

It is important for our project to notice that Mill does not think that desires are the only kind of motivation for action. In the *System of Logic* he writes:

When the will is said to be determined by motives, a motive does not mean always, or solely, the anticipation of a pleasure or of a pain. ... As we proceed in the formation of habits, and become accustomed to will a particular act or a particular course of conduct because it is pleasurable, we at last continue to will it without any reference to its being pleasurable. Although from some change in us or in our circumstances, we have ceased to find any pleasure in the action, or perhaps to anticipate any pleasure in the consequence of

---

<sup>7</sup> See for example Skorupski, 1989, 300.

it, we still continue to desire the action, and consequently to do it. In this manner it is that habits of hurtful excess continue to be practised although they have ceased to be pleasurable; and in this manner also it is that the habit of willing to persevere in the course which he has chosen does not desert the moral hero, even when the reward, however real, which he doubtless receives from the consciousness of well-doing, is anything but an equivalent for the sufferings he undergoes or the wishes he may have to renounce.<sup>8</sup>

The title of this subsection of the *System* is quite telling: “A motive not always the anticipation of a pleasure or a pain.” The motive is still a desire: “we still continue to desire the action.” However, it might be the case that the action no longer yields pleasure or that we cease to anticipate pleasure or pain in the action. That is, even though originally we used to think of the action as pleasant, i.e., we desired that kind of action for its pleasantness, now, we no longer evaluate the action in those terms, the action is no longer conceived as a pleasant action. We no longer attend to our desire to act in that way; we simply, and unreflectively, act in that way. Even more to the point, Mill concludes:

A habit of willing is commonly called a purpose; and among the causes of our volitions, and of the actions which flow from them, must be reckoned not only liking and aversions, but also purposes. It is only when our purposes have become independent of the feelings of pain or pleasure from which they originally took their rise that we are said to have a confirmed character. “A character,” says Novalis, “is a completely fashioned will;” and the will, once so fashioned, may be steady and constant, when the passive susceptibilities of pleasure and pain are greatly weakened or materially changed.<sup>9</sup>

These quotations are clear enough. Even if originally all of our actions originated from motives related to pleasure and pain, that needn't necessarily be the case once habitual willing takes over. Agents capable of habitual willing display volitions that are not caused by the idea of pleasure and the avoidance of pain. Yet Mill says, ‘we still continue to desire the action, and consequently to do it’. This claim, however, seems to be in contradiction with what Mill says concerning desires as necessarily involving the idea of pleasure. In *Utilitarianism*, Mill seems to be addressing this point:

---

<sup>8</sup> Mill, 1936, 551, vi.ii.4.

<sup>9</sup> Mill, 1936, 551-2, vi.ii.4

Will, the active phenomenon, is a different thing from desire, the state of passive sensibility, and though originally an offshoot from it, may in time take root and detach itself from the parent stock; so much so that in the case of an habitual purpose, instead of willing a thing because we desire it, we often desire it only because we will it.... The distinction between will and desire thus understood, is an authentic and highly important psychological fact; but the fact consists solely in this—that will, like all other parts of our constitution, is amenable to habit, and that we may will from habit what we no longer desire for itself, or desire only because we will it.<sup>10</sup>

It seems clear from this paragraph that, leaving aside the role of desires as the ‘parent stock’ of habitual willing, when we act out of habitual willing the object of our desires is no longer ‘desired for itself’. Even though Mill does not tell us what the role of these residual desires might be, one might take Mill to have thought that any action must necessarily spring from a desire. Thus, this desire would be somehow instrumental, a desire to bring about the action which one wills. At any rate, it would not be the same kind of desiring-as-pleasant we reported Mill to describe at the start of this subsection. It would be true, then, that in these cases it is not because an agent finds an object pleasant that he is disposed to pursue that object, but that one is disposed to act irrespective of the pleasurable nature of the object. Mill distinguishes this kind of disposition from that that we normally have when acting out of desires, by saying that this is not a ‘desire for the object itself’. In conclusion, there seem to be two kinds of motives: desires as passive sensibility, i.e., as conceiving of something as good to have under the idea of pleasure (and as we postulated even under other desirability features) and willing something. When we will something we are still moved to act by a desire, but, in this sense, desires are not passive sensibility, there is no idea of pleasure or of any other desirability feature involved. Next, we shall see how these distinctions will help us make our case against Scanlon.

### 2.3. Scanlon’s Millian desires

Similarly to Mill, Scanlon claims that:

---

<sup>10</sup> Mill, 1998, 85, IV.11.



Desiring something involves having the tendency to see something good or desirable about it. ... [H]aving what is generally called a desire involves having a tendency to see something as a reason. Even if this is true, however, this is not all that desire involves. Having a desire to do something (such as to drink a glass of water) is not just a matter of seeing something good about it. I might see something good about drinking a glass of foul-tasting medicine, but would not therefore be said to have a desire to do it, and I can even see that something would be pleasant without, in the normal sense, feeling a desire to do it. Reflection on the differences between these cases leads me to what I call the idea of desire in the directed-attention sense. A person has a desire in the directed-attention sense that P if the thought that P keeps occurring to him or her in a favourable light, that is to say, if the person's attention is directed insistently toward considerations that present themselves as counting in favor of P.<sup>11</sup>

In this passage, Scanlon makes some very important points: we might see something good in doing things we clearly have no desire to do (think about the foul-tasting medicine example) and, sometimes, we might have no desire to do something that would be pleasant, i.e., occasionally, we might have no desire to do things that we normally take to be good. So far, I agree with Scanlon. According to Scanlon, the idea of desires in the directed-attention sense:

...capture[s] an essential element in the intuitive notion of (occurrent) desire. Desire for food, for example, and sexual desires are marked by just this character of directed attention. And this character is generally missing in cases in which we say that a person who does something for a reason nonetheless "has no desire to do it," as when, for example, one must tell a friend some unwelcome news. ...[W]e most commonly speak of "having a desire" in this sense for things that involve the prospect of pleasant experiences (or the avoidance of unpleasant ones). But while it seems to be a fact about us that pleasure and pain are particularly able to attract our attention, they are not the only things that can do so. People can "have a great desire" in the directed-attention sense to succeed in some endeavor, to achieve fame (even posthumous fame), to provide for their children after their death... [I]t is just as appropriate on this account to say that someone with a very active conscience "has a strong desire to do the right thing" as it is to say of a person who is utterly unscrupulous that he "has a strong desire for personal gain".<sup>12</sup>

---

<sup>11</sup> Scanlon, 1998, 38-39.

Once again, I am in agreement with this part of Scanlon's thought. What I would like to show, however, is that this part of his thought is compatible with the Millian distinction between desiring and willing. In Scanlon we find an important distinction: (a) we desire something in the directed-attention sense, i.e., we see that thing as good in some respect; and (b) we may see something good in doing things we have no desire to do. Now, suppose that in the first case an agent rightly takes himself to have reason to desire the thing he desires, i.e., the agent is right in considering that thing to be good in some respect. What are we to say about desires for cases such as (b)? Suppose that even in that case the thing is good in some respect: must the agent have reason to *desire* it? My contention is that there may be no such reason; the agent may simply have reason to do it but no reason to desire it. Why think that agents must have reason to *desire* to do every action they have reason to do?<sup>13</sup>

Nothing we have reported Scanlon as claiming would prevent him from endorsing Mill's distinction between desiring and willing in the following terms: we may have reason to *desire* to do certain actions but we may also simply have reason to do them. In the first case, it would be appropriate for the agent to have a desire to do that thing. In the second case, it would be appropriate for the agent to simply *will* the action that brings about or realises that thing. In both cases, the thing one has reason to do is good in some respect. The question is, however, in what respect are things we have reason to desire good, and in what respect do they differ from things we have reason to will? The idea is simple: an agent has reason to desire all and only things that are good for him or part of his well-being. All other good things an agent has reason to have all sorts of other pro-attitudes towards, including willing. This is a central idea of this thesis and a full defence of it will be developed throughout the chapters to come. In this section, however, I intend to lend initial plausibility to it.

---

<sup>12</sup> Scanlon, 1998, 39.

<sup>13</sup> It might be helpful here to introduce what I take to be three distinct ways of using the concept 'desire' in philosophy. (a) *Substantive or directed-attention desire*: this is the sense of desire we discussed above, to which Mill referred as 'passive sensibility'. (b) *Purpose-eliciting desires*: to say that I want or desire something is to say that the object of my desire is on my agenda. For example, if I knock on someone's office-door it is because I *want* to see this person (that's my purpose). It is a further question to ask whether my desire is substantive as in (a). Finally, (c) *Humean desire*: this is a psychological entity that on a Humean theory of desires always ultimately motivates and justifies our actions.

Consider Mill once again: when an agent desires something he conceives of that thing as pleasant and good in that respect. Clearly, however, the thing is pleasant *for* the agent. Even if the course of action the agent desires is intended to, say, benefit someone else, if the agent desires to do that course of action he will conceive of that course of action as (bringing about something) pleasant for him. As claimed above, however, pleasure should not be considered as the only desirability feature of our ends. Take the testing for AIDS example. I am miserable not knowing because I desire to know. I conceive of the state in which I will come to know the truth about my situation as desirable. But the expressions 'desirable' and 'desirability feature' are not as specific as we would want them to be. Sure enough, I find it desirable to know: but does that mean that I consider it is good *tout court* that I know or, more modestly, that I consider it is good *for me* to know? I believe that the second interpretation is what we really mean in these circumstances. As far as I can see, it might not at all be good *tout court* that I know but that is not what is at stake here. I desire to know because I take it to be good for me to know. It is *desirable for me* to know; nothing is being implied as to whether it is good *tout court*.

It might be retorted that surely we desire things that we take to be good *tout court*. I would agree and add that we would call *virtuous* those agents that do so with unerring constancy. Agents of this type must be contrasted to another type of moral character that we will call *conscientious*. The former *desire* to do that which is good; they conceive of that which is good as also being a part of their good. When an agent acts conscientiously, however, the action is not accompanied by a corresponding desire. The agent must conceive of the action as his duty, not as good or desirable *for him*. The agent will take himself to have reason to will the action though we cannot say that he has reason to desire it.

#### 2.4. The upshot

We now have enough elements to counter Scanlon's argument to the effect that just as we have reason to desire and pursue the various things that contribute to our well-being we have reason to desire and pursue for themselves things that do not contribute to our well-being. *Contra* Scanlon, we individuated desire as the one kind of pro-attitude that all and only prudentially good things would give us reason to have. More precisely: something is *good for* an agent if and only if it has a property that gives reason to the agent to *desire* it. Note that this account (1) is an instance of the buck-passing account of goodness; (2)

allows agents to have reason to do things that do not contribute to their well-being; in those cases, we would say they have reason to *will* those actions; (3) is capable of separating the goods contributing to an agent's well-being from the goods that don't. As we shall see, this account can be endorsed while maintaining that we have reason to want the things that are good for us for their own sake. This will be fully shown in §4 below.

### 3. Butler, self-love, the passions, and happiness

#### 3.1. Why Butler?

We concluded Chapter 2 with the following picture: we take ourselves to have reason to desire and pursue a variety of things for their own sake, i.e., not as a means to something else. Many of these things, however, contribute to our well-being even though that is not what gives us reason to desire and pursue them in the first instance. Yet, we claimed that Scanlon's conclusion to the effect that well-being does not play an (explanatorily) important role in our deliberation does not follow, because this concept is implicitly part of our deliberative process. What we need to get a better grasp of in this chapter is precisely what this concept of well-being is and how it and the things that make it up are related. For these purposes, a discussion of Butler is very much in order as the picture he put forward shares many similarities with the picture we have presented so far.

According to Butler, our *happiness* is realised by the gratification of a number of things that we desire in and for themselves, not because they contribute to our happiness (switch Butler's *happiness* with *well-being* and you will begin to see the resemblance between his picture and ours). Our desires for these things Butler called the particular passions. Furthermore, Butler wanted to say that, though we desire these various things in their own right, we also have a concern for our happiness itself, what he calls self-love, and that it is self-love that should guide the pursuit of the various things we desire. In a vein similar to Butler's, I will conclude my own analysis of well-being by saying that it is the idea of well-being that guides the pursuits of many of the things we take ourselves to have reason to desire for their own sake. We shall limit our discussion of Butler's ethics to a critical

presentation of only some of its fundamental elements. In particular we will discuss three notions, i.e., self-love, the particular passions, and happiness, and the way they are related.

According to Butler self-love is one of the four principles of action<sup>14</sup> that constitute our human nature, the other three being the passions (or affections or appetites);<sup>15</sup> benevolence; and conscience. These elements are arranged in hierarchical order according to the (rational) *authority* of each of them. The passions have the least authority while conscience, a “rational” or “reflective” principle, has the most authority.<sup>16</sup> Self-love and benevolence stand somewhere in between conscience and the passions. In order to understand this ranking we must understand the notion of authority as Butler uses it. The authority of a principle of action is not to be understood in terms of its felt *strength*. The strength with which a certain appetite may press someone into acting may override the more authoritative contrary judgement (‘the disapproval’ in Butler’s terminology) of that person’s conscience. In this circumstance, Butler would say, a man is acting against his nature.<sup>17</sup> The authority of a principle of action, then, is part of its nature and has ultimate normative force. Butler intends self-love to be distinctively more authoritative than the passions.<sup>18</sup>

### 3.2. Self-love, the passions, and happiness: Part I

Butler described self-love as

---

<sup>14</sup> Penelhum, 1985, 26, tells us that the term ‘principle’ is Butler’s “commonest general name for all the motives he discusses, and there does not seem to be any implied contrast between ‘principle’ and ‘instinct’... The term ‘principle’, then, appears to be used indifferently for any conscious inner source of human action.”

<sup>15</sup> Henceforth, I will use the word “passions” to refer to the appetites, the passions, and the affections.

<sup>16</sup> The fact that conscience is reflective and a rational principle did not exclude, for Butler, that “affections” were involved in its workings.

<sup>17</sup> The concept of nature, in Butler, is not a merely descriptive concept, or at least not in the contemporary sense of “descriptive”. According to him, teleology has a clear role in our nature: each principle is to serve a certain function and its authority over the other principles has been established by God as He implanted them in us. It is in this sense that when one acts on a principle which defies the authority of a higher principle, one acts against one’s nature.

<sup>18</sup> Most commentators agree in saying that self-love is a second-order desire, i.e., a desire about one’s desires: Duncan-Smith, 1952. Penelhum, 1985. Henson, 1987, disagrees. Butler, of course, never uses the term ‘second-order desire’.

... the principle of an action proceeding from cool consideration that it will be to my advantage ...<sup>19</sup>

Elsewhere Butler writes:

... men form a general notion of interest, some placing it in one thing, and some in another, and have a considerable regard to it throughout the course of their life, which is owing to self-love.<sup>20</sup>

In other places still, Butler talks of self-love as a person's desire for his known, or manifest and real, interest. There are two interesting features in Butler's discussion: firstly, self-love is a desire or principle of action with regard to one's interest; secondly, this interest will be placed in different things by different people. Let us read more about what these things are and how self-love is "placed" in them.

Every man hath a general desire of his own happiness, and likewise a variety of particular affections, passions and appetites to particular external objects. The former proceeds from or is self-love; and seems inseparable from all sensible creatures who can reflect upon themselves and their own interest or happiness, so as to have that interest an object to their minds... The object the former pursues is somewhat internal, our own happiness, enjoyment, satisfaction; whether we have, or have not, a distinct particular perception what it is, or wherein it consists: the objects of the latter are this or that particular external thing, which the affections tend towards, and of which it hath always a particular idea or perception. The principle we call self-love never seeks anything external for the sake of the thing, but only as a means of happiness or good: particular affections rest in the external things themselves. One belongs to man as a reasonable creature reflecting upon his own interest or happiness. The other, though quite distinct from reason, are as much part of human nature.<sup>21</sup>

Most generally, this passage distinguishes self-love from the passions along three dimensions: (1) self-love is a general principle of action while the passions are particular

---

<sup>19</sup> Butler, 1949, XI, 7, 168-169. I will refer to the W. R. Matthews edition, 1949. This edition, in turn, uses the paragraphing introduced by J.H. Bernard's edition. The reference will be made up by three numbers: in order the (roman) number of the Sermon (except in the case of the Preface), the paragraph number, and the page number.

<sup>20</sup> Butler, 1949, Preface, 37, 22.

principles of action; (2) the object of self-love is internal while the objects of the passions are external; (3) self-love is a rational principle while the passions aren't. In what follows we shall examine what I take to be the most important of these two criteria, namely (1) and (3).

Self-love is an unmotivated desire for a general type of object, i.e., one's own happiness.<sup>22</sup> We saw how Butler defined self-love as the principle of an action done because it would be to one's own advantage, or interest. This is what Butler also calls a *general* desire for one's happiness. This desire is *general* insofar as, at this stage, it is an unarticulated, unmotivated desire, not to be identified with the desire for any specific or more articulated object or state of affairs.

Self-love, or the desire for one's own happiness, can also be articulated more specifically. In order to do so, however, we must introduce the notion of the particular passions. Just as we have an unmotivated general desire for our happiness, we have a variety of *particular* unmotivated passions for *particular* (external) objects. We have a number of different primitive desires each for its naturally suited object (for example, an appetite such as hunger would have food as its naturally suited object). The particular passions are connected to self-love to the extent that happiness arises out of the enjoyment by certain particular passions of their suited object:

...the very idea of interest or happiness consists in this that an appetite or affection enjoys its object.<sup>23</sup>

Happiness or satisfaction consists only in the enjoyment of those objects which are by nature suited to our several particular appetites, passions, and affections. ... [H]appiness consists in the gratification of certain affections, appetites, and passions, with objects which are by nature adapted to them.<sup>24</sup>

Butler seems to be putting forward two theses: according to the first happiness would arise whenever a passion enjoys its object. According to the second, however, happiness would

---

<sup>21</sup> Butler, 1949, XI, 5, 167.

<sup>22</sup> Motivated desires are desires whose force is entirely borrowed from another motive; remove that motive and we would no longer have that desire. Unmotivated desires are primitive in the sense that their motivational force rests completely within themselves.

<sup>23</sup> Butler, 1949, Preface, 37, 22.

arise only when *certain* passions enjoy their objects.<sup>25</sup> Which of these two views does Butler want to defend? The following passage might be taken to support the second, though in this passage Butler is actually putting forward an argument against psychological egoism:

...we should want words to express the difference between the principle of an action proceeding from cool consideration that it will be to my advantage, and an action, suppose of revenge or friendship, by which a man runs upon a certain ruin to do evil or good to another. It is manifest the principles of these actions are totally different, and so want different words to be distinguished by; all that they agree in is that they both proceed from and are done to gratify an inclination in a man's self. But the principle or inclination in one case is self-love, in the other, hatred or love of another. There is then a distinction between the cool principle of self-love or general desire of our happiness, as one part of our nature and one principle of action, and the particular affections toward particular external objects, as another part of our nature and another principle of action.<sup>26</sup>

Butler writes that a "man runs upon a certain ruin to do evil or good to another". When he does, some of his particular inclinations are gratified. What is clear, though, is that it is not self-love which is being gratified since, by definition, that is not the man's motive and, also, because this man knowingly acts so as to incur "a certain ruin". According to Butler, writers such as Hobbes are confused when they call "actions interested which are done in contradiction to the most manifest known interest, merely for the gratification of a present passion".<sup>27</sup> On Butler's view, the simple fact that a passion enjoys its natural object or is gratified is not in and of itself sufficient to claim that the agent's self-love is being served (even though it might still be open to Butler to say that *some* happiness might arise out of it). Interestingly, then, more than just satisfaction of simple desires is required in order for (all things considered) happiness to arise.

The gist of Butler's idea is that self-love is a *general* desire of one's happiness and happiness is constituted by the enjoyment of certain, not all, *particular* passions of their

---

<sup>24</sup> Butler, 1949, XI, 9 and 16, 170 and 178.

<sup>25</sup> Phillips, 2000, 425-6, n. 19, makes a similar point when he writes: "Butler is clearly committed to some subjectivist account of happiness. What is less clear is just what version of subjectivism. Sometimes Butler writes as if happiness is a matter of the satisfaction of prior desires; at other times, the point seems to be the weaker one that things make us happy only if we are constituted so as to enjoy them."

<sup>26</sup> Butler, 1949, XI, 7, 168-169.



suiting object. Someone (and I do have Scanlon in mind) might find it easy to agree with Butler insofar as he claims that there are particular passions. Yet, he may disagree with him insofar as he believes that there is such a thing as a general desire for happiness such as self-love. Butler may try to persuade him in this way:

Suppose another man to go through some laborious work on the promise of a great reward, without any distinct knowledge what the reward will be; this course of action cannot be ascribed to any particular passion.<sup>28</sup>

Let us refer to this example as the ‘laborious work’ example. I find this example rather persuasive in establishing that there is an ultimate or unmotivated desire for happiness that is general and unarticulated. This example shows that what we would find motivating is the abstract idea that my interest would be served, whatever concrete things I place my interest in. Yet, Butler also says,

It is not because we love ourselves that we find delight in such and such objects, but because we have particular affections towards them. Take away these affections, and you leave self-love absolutely nothing to employ itself about.<sup>29</sup>

It is not because we desire our happiness that we find eating delightful (a source of happiness), but because we have a desire for food itself. The point is that more than satisfaction of self-love is needed in order for happiness to arise. But, and this introduces the second difference between self-love and the passions, satisfaction of any passion is not likely to yield happiness unless it is assisted by self-love, the rational principle. According to Butler, in fact, the passions are “blind” occurrences; they can’t ‘see’ further than the satisfaction of their object. They express or embody “a prior suitability between” themselves and their object;<sup>30</sup> they imply “a direct simple tendency towards such and such objects, without distinction of the means by which they are to be obtained.”<sup>31</sup> For Butler the passions are essentially *passive*: God implants them in us and it is a feature

---

<sup>27</sup> Butler, 1949, Preface, 35, 21.

<sup>28</sup> Butler, 1949, I, 7, , 36 n. 5.

<sup>29</sup> Butler, 1949, Preface, 37, 22.

<sup>30</sup> Butler, 1949, XI, 6, 167.

<sup>31</sup> Butler, II, 13, 56.

(independent from our will) that they *naturally* tend towards their object.<sup>32</sup> The suitability between a desire and an object is *prior* in that it is a relation that just happens to exist by nature (or by the will of God) whether or not we intentionally intervene in it to let it run its course. At this level, reason can certainly play an instrumental role: it can help us calculate the best way to satisfy our passions. But even that is not on its own enough to guarantee happiness. Purely instrumental reason might help us choose the best means to satisfy a particular passion that is avowedly against our interest (think about the vengeful man example).

That is why we also need self-love. Of course, self-love too is an affection:

It is an affection to ourselves; a regard to our own interest, happiness and private good: and in the proportion a man hath this, he is interested, a lover of himself.<sup>33</sup>

Were it not an affection, says Penelhum, Butler would have not called it *self-love*.<sup>34</sup> However, it can also be classed as a rational principle. According to Penelhum:

It is a rational principle in the sense that it requires a rational being's capacity to distinguish between such a general and second-order object as happiness and other objects of want, and in the sense that it is an affection that can only be exercised through judgements that particular objectives will, or will not, contribute to this general one. [Butler] brings out this necessary element of calculation by calling it 'cool self-love' and 'reasonable self-love'.<sup>35</sup>

The passions are blind; when acting with an eye to one's "private good", however, we should only act on those particular passions the satisfaction of which is believed to bring about one's happiness. When acting out of self-love, then, reason plays two roles: that of *selecting* which passions to pursue and which not to pursue and the instrumental role we mentioned above.<sup>36</sup> In other words, self-love requires rationality about the ends, as well as rationality about the means. Borrowing Butler's own example, a man can see that he

---

<sup>32</sup> See Penelhum, 1985, 29.

<sup>33</sup> Butler, XI, 8, 169.

<sup>34</sup> Penelhum, 1985, 29.

<sup>35</sup> Penelhum, 1985, 30.

<sup>36</sup> On this point see McNaughton, 1992, 280-281.

desires to kill another man for revenge but he can also see that that would bring certain ruin upon him; out of self-love he can see he should not act on this vengeful motive.

### 3.3. Self-love, the passions, and happiness: Part II

Butler insists on distinguishing self-love from its object, i.e., happiness:

Happiness does not consist in self-love. The desire for happiness is no more the thing in itself, than the desire of riches is the possession or enjoyment of them. People may love themselves with the most entire and unbounded affection, and yet be extremely miserable. ... if self-love wholly engrosses us, and leaves no room for any other principle, there can be absolutely no such thing as happiness, or enjoyment of any kind whatever; since happiness consists in the gratification of particular passions, which supposes the having of them. Self-love does not constitute *this* or *that* to be our interest or good; but our interest or good being constituted by nature and supposed, self-love only puts us upon obtaining and securing it.<sup>37</sup>

A person's happiness should not be confused with this person's desire for his happiness, i.e., happiness and self-love are distinct things. Remarkably, in this passage Butler also offers an early formulation of the paradox of hedonism: a person's desire for his happiness may be so strong as to impede him from being happy. Often, when this situation occurs, the agent no longer sees that happiness arises from the satisfaction of certain *particular* passions. Which passions? Part of the reply is already mentioned in the quotation above: "our interest or good being constituted by nature". Another part of the reply comes a few paragraphs below in Sermon XI, 15-17. Here, Butler examines one by one the objects of some of our particular passions. He begins with the enjoyment of those objects, "which are by nature adapted to our several faculties". They are supposed to arise from riches, honours, and the gratification of sensual appetites. But, then, he goes on:

... men of all ages professed to have found satisfaction in the exercise of charity, in the love of their neighbour, in endeavouring to promote the happiness of all they had to do

---

<sup>37</sup> Butler, 1949, XI, 9, 170.

with, and in the pursuit of what is just and right and good, as the general bent of their mind, and end of their life.<sup>38</sup>

Butler goes as far as considering the gratification certain people would derive from “approving themselves to God’s unerring judgement, to whom they thus refer all their actions”.<sup>39</sup> Even the gratification of that kind of passion may be part of a person’s happiness. Butler, however, goes to a certain length to insist that:

Self-love may indeed set us on work to gratify these [affections, appetites, passions]; but happiness or enjoyment has no immediate connection with self-love, but arises from such gratification alone. ... they [the affections, appetites, and passions] equally proceed from or do not proceed from self-love, they equally include or equally exclude this principle.<sup>40</sup>

Happiness and self-love are two distinct things. What constitutes a person’s happiness depends on his or her nature even though the kinds of things from which happiness may arise seem to be limited by our nature as human beings and by facts about each particular person. Some of us may be satisfied simply with honour, riches and sensual pleasures. Others may find satisfaction in other endeavours, such as promoting the good of others, and even in acting in accordance with the will of God. As for self-love, it “only puts us upon obtaining and securing” the various things from which our happiness arises. Interestingly, the impulse to satisfy the passions from which happiness might arise may or may not proceed from self-love. Take benevolence to be the particular passion to further another person’s good. One might “set to gratify benevolence” from self-love or, maybe, from the particular benevolent passion itself. In the first case, I suppose, to act benevolently might involve the realisation that one’s desire to help another is part of one’s good, interest, or happiness.

Let’s sum up Butler’s view. We do have a regard for our own interest throughout our lives but whenever we are to realise this interest we need to act on some of our particular passions; “Take away these affections, and you leave self-love absolutely nothing to employ itself about.” That is because the object of our self-love is happiness and happiness

---

<sup>38</sup> Butler, XI, 15, 176.

<sup>39</sup> Butler, XI, 15, 177.

<sup>40</sup> Butler, XI, 16, 178.

arises when some of our particular passions are gratified. Notice that the relation between self-love, the particular passions and happiness play a role at two different levels, a psychological one and an ontological one. At a psychological level we desire both our happiness in general as well as the various objects of the particular passions. From an ontological point of view, then, the states in which my happiness can arise are only the states in which some particular desires that I have are satisfied. Happiness, however, does not arise through the gratification of any of the particular passions that I happen to have. Through self-love, reason informs our selection of the particular desires we should act on in order to enhance our chances to see our happiness realised. Butler's example of the vengeful man is a good illustration. Only our rationally informed desire for happiness is capable of guiding our action towards what is in our interest by choosing to act on only some of our particular desires.<sup>41</sup>

#### 4. The idea, the parts, and the sources of well-being

With this Butlerian introduction in mind, in this section, I shall put forward an analysis of the concept of well-being. To anticipate a little, we shall see that the concept of well-being involves (1) the *idea* of well-being; (2) the parts of well-being; and (3) the sources of well-being. I shall take each of these notions in turn and examine the way in which they are related to each other. Finally, I shall note that both rational and non-rational creatures may be capable of well-being though what the concept of well-being should be understood as involving when applied to either kind of being is liable to change.

---

<sup>41</sup> To the contemporary reader, Butler may seem confused on at least one count. He writes (1) that happiness is or consists in certain particular passions' *enjoyment* or *gratification* of their naturally suited objects; and (2) that happiness consists in (a) the *enjoyment* by a man of honour, riches and sensual pleasures, or (b) the *satisfaction* of some men in the exercise of charity and the love of others, or (c) the *gratification* derived by some men in acting in accordance to the will of God. There seems to be some shift between the two senses of satisfaction, enjoyment, or gratification expressed by (1) and (2). (1) seems to point out to what today we would call the logical sense in which *a desire is satisfied* when the state of affairs that appears as its content obtains. (2) is the sense in which satisfaction, gratification, and enjoyment refer to the *positive feeling* that arises when one of the agent's desires obtain. Today, we would want to associate the first type of enjoyment to desire-satisfaction theories of happiness or well-being and the second type of enjoyment with some form of hedonism or with *mental statism* (a view according to which a person's happiness or well-being can be affected only by certain kinds of mental states). We shall not discuss from an exegetical point of view which of these options Butler would have to endorse. Nor shall we examine the relation between desire-satisfaction

#### 4.1. The idea of well-being

Suppose, with Scanlon, that we have reason to desire enjoyment for its own sake. Suppose that an agent is confronted with a choice between two courses of action with no consequences for anyone other than himself. Suppose also that the only desirability feature of each action is its enjoyableness. We should agree that it would be most rational for the agent to choose the thing that brought about the most enjoyment. What makes that the most rational choice? One answer would be 'enjoyment itself' given that we are postulating that we have reason to desire it for its own sake. But then we could carry out a simple test we already introduced at the end of the last chapter. Suppose it could be shown that the most enjoyable thing was not in fact good for the agent. In other words, suppose that this thing was not part of the agent's well-being: would we still think that he has most reason to want it? Clearly not, at least not in a case like the one we are considering. What confers rationality to our desire for that thing, then, is not simply enjoyment. Rather, it seems that the idea of the agent's well-being determines what there is most reason to desire.

Let me run a slightly different version of this argument. We agreed that it would be most rational for the agent to desire the thing that brought about the most enjoyment. If someone asked why that would be the most rational thing to desire and pursue one could answer that that is so simply because it is believed to yield the most enjoyment. Scanlon would have to be satisfied with this answer. But wouldn't we find this answer satisfactory only insofar as we also took it to imply that that thing was the one most in the agent's interest? If someone made us seriously doubt the fact that this thing is the one most in the agent's interest, wouldn't we *ipso facto* doubt that the agent has most reason to desire it?

Next, let us alter the description of our example in the following way. We can no longer assume that the consequences of the agent's choice are restricted to himself alone. Other persons might be involved. We also know that the choice resulting in the most enjoyment for the agent is the best in terms of the agent's well-being and, yet, he chooses something else. What are we to conclude? I think we have three options: (a) the agent has acted on considerations other than his well-being; (b) the agent did act on what he believed to be his interest but held false beliefs concerning what was the best choice for him. Suppose,

---

theories of well-being and hedonism/mental statism. An excellent discussion of this latter topic can be found

however, that his choice cannot be explained by either (a) or (b), then, we will have to conclude that (c) the agent acted irrationally. In these circumstances, the agent's rationality is criticisable on the grounds that he did not take the option that he had good reason to believe to be most conducive to his well-being. If this conclusion is correct, then we can claim that the idea of one's well-being is a normative idea. More precisely, as we saw in §2, the idea of well-being entails reasons to desire things for the agent whose well-being it is. Now, as we shall explain in Chapter 8, when we have reason to desire something we have reason to do what would realise the object of our desire.

We have been using the expression of the 'idea or concept of well-being' and claimed that it entails reasons; but what is this *idea* of well-being exactly? By 'the idea of well-being' I mean precisely the same as Butler's unarticulated "general notion of interest" "advantage", "private good", or benefit, "some placing in one thing, and some in another", and for which we have a considerable regard throughout the course of our life. I need say no more about it and I am confident the readers will know what I am referring to.

#### 4.2. The parts and the sources of well-being

The idea of well-being stands in a particular relation to the parts of well-being. Butler thought that self-love would have nothing to employ itself about without the particular passions and that happiness could not arise without them. Similarly, I think that a person's well-being could not arise without that person's attaining the various things that he has reason to desire for their own sake. These things I will call the parts, components, constituents, or ingredients of well-being. When a person's well-being is said to arise, in substantive terms, we mean to say that one or another of the components (or a certain arrangement of components) of this person's well-being is being instantiated in his life. If enjoyment is one such component, then its instantiation in the agent's life fleshes out or gives substance to this person's well-being.

Or again, suppose that on one occasion, say, an enjoyment and a particular achievement were both part of the agent's well-being and came to conflict with each other: he can pursue one or the other but not both. In other words, he has reason to desire for their own

---

in Kagan, 1992.

sake both things taken individually though, unfortunately, pursuit of the one excludes the pursuit of the other. After brief deliberation, the agent comes to the (right) conclusion that on this occasion he has most reason to desire enjoyment. In this case, the outcome in which the agent obtained more enjoyment and less achievement is better for him than the outcome in which he obtained less enjoyment and more achievement. Though it is true of this outcome that it contains more well-being that should not be understood as meaning that this outcome contains anything other than a certain arrangement of enjoyment and achievement. In substantive terms, a person's well-being is nothing over and above (a particular arrangement of) its parts. This thought expresses my own view that well-being is not a separate but an inclusive good.<sup>42</sup>

It should be emphasised, however, that in non-substantive terms the components of well-being are not all there is to well-being. We have explored what we mean by the *idea* of well-being in the subsection above: it is the general, unarticulated idea of something being in a person's interest, for his good or advantage or benefit. Remember how Butler found self-love to be motivating (think about his "laborious work" example). Just as Butler, and contrary (or largely contrary) to what Scanlon claims, I too believe that the idea of well-being is motivationally active though many times it is so only implicitly. True enough, we are primarily motivated by the various things that make up our well-being: pleasure or enjoyment, and whatever other desirability features we take our ends to have. But both the motivational and justificatory power of these features is held in check by the idea of well-being. Take that idea away, i.e., show someone that desiring and pursuing these things would amount to no good for him, and away goes much of their power. Implicitly or explicitly, the idea of well-being is explanatorily and justificatorily important.

---

<sup>42</sup> Griffin seems to share a similar point of view when he writes: "So when, for whatever purposes, we shift from everyday talk of pursuing various different ends to theoretical talk of maximising a single quantity, 'utility', this quantity should not be understood as an end of the same kind, only grander. There is simply no case for reducing these various ends to a single end in this sense... When our various values conflict, we may attempt to resolve the conflict by trying to realize as much 'value' as possible...even in the absence of one substantive end as mediator." Or again: "...when we assess various combinations of our ends, there is no further substantive value, no supreme or dominant end, to rank them." And finally: "Since utility is not a substantive value at all, we have to give up the idea that our various particular ends are valuable only because they *cause, produce, bring about, are sources of*, utility. On the contrary, they are the values, utility is not." Griffin, 1986, 32, 36, and 332 n. 24. Of course, I would not go as far as Griffin's last claim. Well-being is a value, namely, that which is good for one. Substantively, however, this value is realised only through its parts or components.



Next, think back to Irwin's view concerning the parts of *eudaimonia*. Taking friendship to be something we have reason to desire for its own sake, Irwin took it that there is one single feature  $F$  with two descriptions:  $F$  is the feature that makes friendship worthwhile;  $F$  is numerically identical to the feature that makes friendship a part of happiness. Within our modified version of the buck-passing account of goodness, we would say that friendship is good for an agent if, and only if, friendship has a property  $F$  such that  $F$  gives reason to desire friendship. Contrary to what Irwin seems to imply, on our account friendship is not simply worthwhile, but it is worthwhile in a certain respect, i.e., as a part of the agent's well-being. This does not imply that friendship may not be good in other respects. Maybe we have reason to admire instantiations of the purest and highest forms of friendship. The question here, however, is whether we have reason to *desire* it for ourselves, and if we do, then, on our account, friendship is part of our good.

Finally, let's clarify the distinction between the parts of well-being and its sources. Suppose that enjoyment or pleasure was the only component (or part, or ingredient) of my well-being. Suppose now that I find eating a chocolate bar pleasurable. In this case, the chocolate bar is the *source* of my well-being. The chocolate bar contributes to my well-being insofar as I find eating it pleasurable. The property of being pleasurable for me is not intrinsic to the chocolate bar. The relevant property intrinsic to the chocolate bar is the property of tasting like chocolate. It then just so happens that this property instantiated by the chocolate bar causes pleasure to arise in me when I eat it. What is good for me is the state in which pleasure arises due to my eating of the chocolate bar. The chocolate bar is a source of my well-being insofar as it contributes to the coming about of this state. This state, however, is a part or component of my well-being not its source.

#### 4.3. Well-being: both normative and descriptive

From this discussion we ought to remark that a cluster of closely connected concepts are involved in our linguistic usage of the concept of well-being. There is the idea of well-being, the parts and the sources of well-being. Next, think about the following judgement: my favourite plant is not doing as well now as it was doing last month. Clearly, this is a comparative well-being judgement. My judgement will probably be given on certain grounds specific to plants in general. I may for example have remarked that the leaves are falling and those that remain are not quite as green. I decide to get my plant back to her

best shape. So, I ask myself what things make plants such as this one fare well. The answer is a certain quantity of water per week and a certain quantity of exposition to sunlight every day. Now, we could say that water and sunlight are the sources of my plant's well-being. After a few weeks, the plant is green with new leaves. The plant is doing well again; we can assume it gets the right amount of the things that are good for it. Its 'doing well' is the state in which the plant enjoys the right combination of water and sunlight. More generally, we could say that well-being is intended as *the state in which a being attains a certain combination of things that are good for it*. Greater well-being is intended as the state in which a being attains a better combination of the things that are good for it. Similarly for lesser well-being.

It is important to notice here that when we make a comparative well-being judgement as we did above we are making a *descriptive* kind of claim. 'The plant did better at time *t*' does not seem to imply any reason for the plant (though maybe that judgement may imply reason to act for those who are concerned with the well-being of this plant). Things, however, are quite different from the point of view of rational creatures. Both rational and non-rational creatures are capable of well-being, but this fact should not let us overlook an important difference in the way the concept of well-being applies to these two different types of being. In the case of rational creatures, the fact that a state is described as one containing greater well-being can motivate and justify the desires and actions of the rational being whose well-being it is. A claim expressing this fact would be a normative claim for that agent. However, only beings with the *linguistic* and *intellectual* capacity to understand the meaning of the *idea* of well-being, i.e., the idea that something is good for one, or in one's interest, advantage or benefit, would find this fact motivating and justifying. More precisely, only beings of this kind *could* be motivated by the description of a state as one containing a certain quantity of their well-being. To say that they *could* be motivated by this fact means that they have the capacity to be so motivated. Yet, beings with both capacities may on some occasion fail to be motivated to desire and pursue something that they recognise is good to have (in their interest, etc.). If they did so systematically, though, we would begin to doubt the claim that they understand the meaning of these terms.

The fact that we as rational beings can describe the conditions under which certain non-rational creatures would do well should not lead us to conclude that the notion of well-

being that applies to us is the same as the notion of well-being that applies to them. These non-rational beings do not have the capacity to represent to themselves the state in which some measure of their well-being obtains in these precise terms. Of course, they may naturally be inclined to pursue that state but not under that description or representation. The plant may struggle to reach its optimal conditions but not because it understands those conditions would be good for it. Or again, given a choice, a mouse may choose to go for cheese rather than lettuce. That, however, is not because it could in principle recognise that that state could be described as what is best for it. The only sense of well-being that we can safely ascribe to non-rational beings is the descriptive notion of well-being. That notion, however, will not be sufficient to elucidate the full meaning of the concept as it applies to rational creatures. The concept of well-being with which we are concerned here, is one that applies only to beings that *could* describe their actions as aiming at a state containing a certain quantity of their well-being.

Importantly, the actions need not be described in those terms but must be capable of bearing that description. It is *not* my view that each and every action affecting the agent's well-being on which the agent is motivated to act *must* be described by the agent in terms of its impact on the agent's well-being. That is, if I am asked *why* on one occasion I choose to go to the cinema, I do not have to describe my action as aiming at a state containing more well-being. As Scanlon would want to say, I might take myself to have reason to go to the cinema simply because I enjoy going to the cinema. Yet, we would want to add that if the action *could* be described as an action affecting my well-being negatively and, being aware of that fact I nonetheless decide to go, either I am acting against my better judgement or I am acting on considerations other than my well-being.

## 5. Conclusion

We now need to bring the conclusions reached so far to bear on our initial discussion. It might be best to bring back to our attention Scanlon's main contention against the role of the concept of well-being:

... this concept of well-being has surprisingly little role to play in the thinking of the rational individual whose life is in question. It sounds absurd to say that individuals have no reason to be concerned with their own well-being, because this seems to imply that they have no reason to be concerned with those things that make their lives better. Clearly they do have reason to be concerned with these things. But in regard to their own lives they have little need to use the concept of well-being itself, either in giving justifications or in drawing distinctions. In particular, individuals have no need for a theory that would clarify the boundaries of their own well-being and provide the basis for sharper quantitative comparisons.<sup>43</sup>

We are in agreement with Scanlon insofar as many times we need not make an *explicit* appeal to the idea of well-being in order to explain and justify our pursuits. But that is far from being sufficient evidence to show that the idea of well-being does not play an important role in our first-personal deliberation. I have been illustrating the way in which implicit knowledge of whether something is or is not good for one may explain away the apparent absence of the concept of well-being in our deliberation. In order to prove that, I showed how explanatorily important the *idea* of well-being in fact is. An agent is justified in taking himself to have reason to desire something to the extent to which this thing is good for him. To say this is precisely to show that there exists a relation between the various parts of well-being, i.e., the various things we take ourselves to have reason to desire for their own sake, and the idea of well-being, i.e., the idea that something is in one's interest, or to one's benefit or advantage.

These conclusions relied on the results we reached in the preceding sections. Here, I ought simply to acknowledge my intellectual debt to Butler and Mill. The former provided the model on which I based my account of the relation between the idea and the parts of well-being. The latter provided the distinction between desiring and willing. This distinction was essential in amending the buck-passing account so as to maintain another distinction, namely that between things that contribute to an agent's well-being and things that don't.

Finally, consider again Scanlon's two criteria of importance of the concept of well-being. Interestingly enough, we find ourselves in agreement with (C1) and (C2). Take pleasure to be an end in itself and part of a person's well-being. Now consider (C1): do we take

---

<sup>43</sup> Scanlon, 1998, 109-110.

pleasure to be important because we take it to contribute to well-being? There is a clear sense in which we do: remove the idea of well-being and pleasure ceases to be considered as what a person has *most* reason to desire and pursue. The answer is the same for (C2): does it make an important difference to know that the pursuit of one's pleasure does not contribute to one's well-being? Surely it does: tell yourself that a certain pleasurable end is not most in your interest, and you should cease to take yourself to have most reason to desire it.

## Part II                    The Theory, the Parts, and the Limits of Well-Being

### Chapter 4                    A Theory of Well-Being

#### 1. Introduction

In this chapter, I will defend an account of well-being whose predecessors might be considered to be J. S. Mill and Sidgwick. On this view, well-being is characterised as what an agent has reason to desire.<sup>1</sup> More precisely, given that my view is a version of the buck-passing account, we should say that something is good *for* an agent if and only if it has a property that gives reason to the agent to *desire* it. I will defend this theory by examining the role desires play in determining an individual's good. In a nutshell, I will argue that though desires play a very important *epistemic* role (§2) they do not semantically determine the good for an agent, not even under idealised conditions (§3). My view then must not be confused with another popular theory according to which a person's well-being or good is defined as what that person would desire under ideal conditions. The final section of this chapter will spell out the difference between our reason-to-desire view and ideal preference type of views (§4).

#### 2. The epistemic role of desires

##### 2.1. Reasons

---

<sup>1</sup> As we shall see in the next section, Mill can only be taken to be a very remote predecessor of this account. In §4.1., I shall claim that Sidgwick clearly put forward a version of the reason-to-desire account though scholarly opinion is rather divided on this point. More recently Skorupski, 1999a, Chapters 2 and 6; 1999b; 2002; and Sayre-McCord, 2001, have developed similar accounts.

On my view, well-being is characterised as what there is a reason for an agent to desire.<sup>2</sup> It is important, then, to begin by saying something about reasons in general. Reasons are relational facts divided in three categories: reasons to believe, reasons to feel, and reasons to act.<sup>3</sup> I will call these types of reasons respectively, epistemic, affective, and practical reasons and give the following analysis of reasons: Reasons are sets of facts standing in a certain relation to an agent and an act-type (a belief, a feeling, or an action). Reasons also have a degree of strength and a time. Thus, for example, were  $\phi$ -ing to give  $x$  pleasure, that would be a reason (of degree  $d$  at time  $t$ ) for  $x$  to  $\phi$ . For example, were smoking to give Jim pleasure that would be a reason for Jim to smoke. Thus, the reason relation in this case holds between an agent, Jim, an act-type, smoking, and the fact that Jim would find smoking pleasant. Desires are a kind of affection. It will follow then that propositions about well-being, i.e., claims about what there is reason for an agent to desire, will belong to the wider category of affective normative propositions.<sup>4</sup> In what follows, I will illustrate the way in which our desires function as the primitive criteria in judgements about well-being.

## 2.2. Desires as the primary criteria for normative judgements about well-being

In Chapter 2, I endorsed a Millian view according to which having a desire for something involves the tendency to see that thing as desirable, worthy of desire, or good to have (under the idea of it as being pleasant or under some other desirability feature), a view which was very similar to Scanlon's conception of desires in the directed-attention sense. As we shall see, Mill is a source of inspiration even for our present task, i.e., showing the epistemic roles of desires in our judgements about well-being. Mill famously argued that:

The only proof capable of being given that an object is visible, is that people actually see it.  
The only proof that a sound is audible, is that people hear it: and so of the other sources of

---

<sup>2</sup> Henceforth, I will use this shorter formulation to express the view that something is good *for* an agent if and only if it has a property that gives reason to the agent to *desire* it.

<sup>3</sup> Unless specified otherwise, our discussion will be conducted in terms of *pro tanto* reasons and *not sufficient* reasons.

<sup>4</sup> The connection between affective and practical reasons will be examined in greater detail in Chapter 8. It suffices here to say that if there is reason for an agent to desire to do something there is (*pro tanto*) reason to do that thing.

our experience. In like manner, I apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.<sup>5</sup>

We take it that Mill does not hereby intend to give a proof in the deductive sense.<sup>6</sup> Mill believed that first principles were not capable of proof by reasoning. That is true of both the principles of knowledge and the principles of conduct. However, our senses and our consciousness provide us support or *evidence* for such principles. Mill's analogy between the epistemic roles of the senses and desires is explained as follows by Geoffrey Sayre-McCord:

Mill's suggestion is that, when it comes to the first principles of our conduct, desires play the same epistemic role that the senses play when it comes to the first principles of knowledge... [T]he evidence we have for our judgments concerning sensible qualities traces back to what is sensed, to the content of our sense-experience. Likewise, ...the evidence that we have for our judgments concerning value traces back to what is desired, to the content of our desire.<sup>7</sup>

Thus, McCord goes on to say, when we are looking at something that appears red, we have overrideable and defeasible evidence that the thing is red. Also, if things never looked red to us we could never get evidence that things were red, nor would we have developed the concept of redness.

Similarly, when we are desiring things, when what we are considering appears good to us, we have evidence (albeit overrideable and defeasible) that the thing is good. Moreover, if we never desired things, we could never get evidence that things were good, and would indeed have never developed the concept of value.<sup>8</sup>

The parallel between sense perception and desiring is illuminating but does not work quite all the way through. This can be seen with the help of something Kant said:

---

<sup>5</sup> Mill, 1998, 81, IV3.

<sup>6</sup> See for example McCord, 2001, 338. That is true whatever one takes Mill's thought to be concerning the legitimacy of deductively valid arguments as proofs.

<sup>7</sup> Sayre-McCord, 2001, 338.

<sup>8</sup> Sayre-McCord, 2001, 339.



Our nature is so constituted that our *intuition* can never be other than sensible; that is it contains only the mode in which we are affected by objects. The faculty, on the other hand, which enables us to *think* the object of sensible intuition is the understanding. To neither of these powers may a preference be given over the other. Without sensibility no object would be given to us, and without understanding no one would be thought. Thoughts without content are empty, intuitions without concepts are blind.<sup>9</sup>

In the first Critique, Kant calls these two powers or capacities of the mind respectively *receptivity* and *spontaneity*, or again *sensibility* and *understanding*. In the remaining part of the paragraph quoted above, Kant continues by claiming that these two capacities cannot exchange their functions. Though knowledge can only arise through their union, their respective contributions should not be confused. On the contrary, one should separate the two carefully and distinguish the science of the rules of sensibility (general aesthetic) from the science of the rules of the understanding, which Kant calls general logic. Now, general logic in its pure form, Kant believes, deals only with *a priori* principles. Contrary to Kant, for our purposes it suffices to say that the principles issuing from spontaneity are *warranted a priori* but defeasible.

Let us go back to Mill. In the case of perception, receptivity and spontaneity are distinguished from but interacting with each other. There is a sensible input and a spontaneous response to it. Without the contribution of the two powers, no judgement would be possible, no thought of anything being red would be possible. Supposing this picture to be true, can a parallel really be drawn with the case of a desire for *x* and the judgement that *x* is good to have? The parallel does not work all the way, as, in the case of desiring, nothing can play a role comparable to the role played by receptivity in the case of sense perception. Desiring something involves no receptivity, no response to a sensible 'intuition'. On the one hand, under normal conditions, the judgement that *x* is visible (or red and visible) is stimulated by *x*. In this case, we must assume that receptivity is activated or stimulated by something, or else we have to postulate that we are subject to some form of hallucination or delusion (i.e., we would not be under normal conditions).

On the other hand, when I judge that *x* is desirable, my judgement is not stimulated *in the same way* by the object of my desire. If we believed that one's receptivity is stimulated in

---

<sup>9</sup> Kant, 1968, A51/B75, 93.

this case as well, we would have to postulate that we could sensibly intuit desirableness along with sounds, colours, odours, etc. But wouldn't that entail the existence of a sixth sense? Also what sensible thing would one be in contact with if one desired *now* a future state of affairs, say, *that I have sweet dreams tonight?* The 'passive sensibility' of desire does not require receptivity and involves no assumption that I am in contact with any worldly desirability feature. To make this assumption would involve that agents are directly in contact with an intrinsically desirable feature, and that what impinges on their receptivity is precisely this desirability. This type of feature would have to have good-to-haveness (or, as J. L. Mackie would say, to-be-pursuedness) built into it and it would be queer to think that we are causally in contact with such normative entities.<sup>10</sup>

If it is agreed that receptivity does not play any role in desire as experience or 'passive sensibility', then, desires must belong to the realm of *pure spontaneity*. I will explore this possibility soon, not before having recapitulated my exposition so far. Mill takes it that an analogy can be drawn between the respective epistemic roles played by the senses in establishing the first principles of knowledge and the desires ('as passive sensibility') in establishing the first principles of conduct. I then inquired into the validity of this analogy by comparing the nature of our judgements involving the senses with our judgements involving desires. Aided by some Kantian terminology, I concluded that there is at least one important difference between judgements about the world that involve the senses and judgements about desirability that involve desiring.

That, however, does not mean that there is no analogy at all between the epistemic roles of the two. The fact that I seem to see an apple on the table goes some way towards establishing that there is an apple on the table. Similarly, the fact that I desire an apple goes some way towards establishing that the apple is desirable or good to have.<sup>11</sup> To put the same analogy in terms of reasons: the fact that I seem to see an apple on the table gives me reason to believe that there is an apple on the table. Similarly, the fact that I desire an apple gives me reason to think that the apple is desirable. In both cases, one's disposition to

---

<sup>10</sup> However, I disagree with Mackie's next step, i.e., the conclusion that no objective values are true. See Mackie, 1977, Chapter 1.

<sup>11</sup> It should be pointed out here that these analogies are not exactly the ones put forward by Mill ("The only proof capable of being given that an object is visible, is that people actually see it. The only proof that a sound is audible, is that people hear it: and so of the other sources of our experience. In like manner, I

desire something or to believe something is accompanied by a corresponding disposition to find that thing desirable or credible.

In general terms: an agent's spontaneous disposition to  $\phi$  (believe, act, feel) is the *primitive criterion* for judgements of what one has reason to  $\phi$ . Thus, for example, the primitive criterion for the judgement that there is reason to admire  $x$  is the feeling of admiration for  $x$ . Or again, the fact that I desire  $g$  typically carries with it a normative impulse, namely the evaluation or judgement that there is reason for me to desire  $g$ .<sup>12</sup> In conclusion, desires, and affections in general, constitute the primary *epistemic* criteria for normative affective judgements. What we should try to understand next is under what circumstances an agent's primitive defeasible judgement that she has a reason to desire something can yield a warranted judgement to that effect.

In a nutshell warranted judgements about well-being will typically start from the disposition to see something as desirable. From there, however, we ought to ask two questions: is one's spontaneous disposition genuine? And, what kinds of defeaters are relevant to claims of the form ' $x$  has reason to desire  $y$ '? I shall not develop here my own answers to these questions so much as rely on conclusions already fully worked out.<sup>13</sup> A spontaneous disposition is genuine when it is experienced by the agent as 'his'. That, for example, would not be the case when the agent is disposed to desire something without being disposed to see that thing as desirable. Also, a disposition to see something as desirable would not be genuine if it arises as the result of indoctrination. Yet, when one's genuine spontaneous disposition is wrong, it should normally be possible to 'educate' one's spontaneity, being careful not to turn this process into a process of indoctrination. The main defeater of our judgements about well-being is the disagreement of competent others in rational discussion. Warrant may come at different levels of authority. Our warranted judgements may be defeated by the disagreement of more authoritative judges.

---

apprehend, the sole evidence it is possible to produce that anything is desirable, is that people do actually desire it.")

<sup>12</sup> This evaluation, however, does not *ipso facto* make it case that  $g$  is desirable. In the words of Hume, to allow that step would amount to *deducing* an 'ought' from an 'is'. I believe that to be an invalid step. I also doubt that Mill fell into this trap, as G. E. Moore famously suggested. Many readers of Mill jump to that conclusion from the infelicitous wording of Mill's proof together with his remark that 'to desire something is the same phenomenon as thinking that thing to be good'. Notice, however, that the claim 'desiring  $x$  is the same phenomenon as thinking  $x$  to be desirable' does not commit Mill or anyone else to the claim that ' $x$  is desirable if and only if  $x$  is desired'. It commits Mill only to saying what we *think* to be desirable. And that is all very well.

Convergence with the authoritative judgements is not what makes judgements true, however. Rather, it is a regulatory ideal, an indicator that one's judgements are on the right track. Finally, this view does not exclude the possibility that on certain normative issues there will always be an irresolvable fracture in the opinions of authoritative judges.

### 2.3. Ex ante and ex post desires

How do we know whether there is reason for someone to desire something? Experience and reflection teach us that we have unmotivated desires for very many objects and states of affairs and that there are several types of desirability features under which we desire these objects and states of affairs. We may form intrinsic desires for various particular objects or states of affairs. Surely, reflection and experience-based agreement of dispositions shows that there are other desirability features beside pleasure. In Chapter 6, we will discuss some of these desirability features in some detail.

Let us suppose that *D* is one such desirability feature. The question we should ask now is: how can we tell that there is reason for an agent to desire object *o* as a source of *D*? We saw how the primary criterion for something being desirable is whether we are disposed to desire it. Suppose that I have a desire for an object *o* under the idea of *D*. The question then is: is this a genuinely spontaneous desire? Suppose that it is and that it is accompanied by my disposition to take myself to have reason to desire *o*. Now, the question is, is there reason to desire *o*? Is *o* desirable? In order to find an answer I might attempt to gather information about *o* and see whether the judgements of people who are well acquainted with it converge in finding *o* desirable.

So far, we have considered spontaneous desires as primitive criteria for judgements about well-being as they occur *before* the object of one's desire is obtained. We will call these desires *ex ante* desires. However, desires may also be useful as *ex post* criteria. One common way to know about what things are good for one is reflecting on the spontaneous affections one feels once the putative good has already been obtained. Rather than being a desire for the obtaining of a certain object or state affairs, one wants one's situation to remain unchanged with respect to what one believes to be a certain feature and in virtue of

---

<sup>13</sup> See Skorupski, 1999a, 1999b, 2002.

that feature. When one is disposed to have an *ex post* desire for something and the accompanying disposition to take oneself to have reason to desire it, one should examine the spontaneity of the desire and the relevant opinion of competent judges. *Ex post* desires in the sense just described are also referred to as *endorsements*.

Note that *ex post* desires can be associated with the attitude of ‘being happy *with*’ something. Being happy or satisfied *with* something is different from being happy in the experiential sense entailed by hedonists. It is also different from what we may call the *feeling* of happiness where that involves joyfulness, high-spiritedness and similar feelings.<sup>14</sup> When one is happy *with D* one wants one’s situation to remain unchanged with respect to what one believes to be a certain feature and in virtue of that feature. However, being happy with something is not necessarily linked to one’s well-being. I may for example say “I am happy with the way the elections went” without necessarily thinking that the result of the elections is good for me. Alternatively, we may be happy with our lives in general, as well as with particular parts of it. Wayne Sumner, for one, thinks that:

Being happy in this sense means having a certain kind of positive attitude toward your life, which in its fullest form has both a cognitive and an affective component. The cognitive aspect of happiness consists in a positive evaluation of the conditions of your life, a judgement that, at least on balance, it measures up favourably against your standards or expectations. This evaluation may be global, covering all the important sectors of your life, or it may focus on one in particular (your work, say, or your family). In either case it represents an affirmation or endorsement of (some or all of) the conditions or circumstances of your life, a judgement that, on balance and taking everything into account, your life is going well for you. ... However, there is more involved in being happy than being disposed to think that your life is going (or has gone) well. The affective side of happiness consists in what we commonly call a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it.<sup>15</sup>

In Sumner, we can see both the affective side of happiness, i.e., “a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it” and a cognitive side, a *positive evaluation* that one’s life or some aspect of it is measuring up

---

<sup>14</sup> In Chapter 6 I discuss the distinction between the sense of ‘happiness’ used by hedonists such as Bentham, Mill, and Sidgwick, and ‘happiness’ as joyfulness or high-spiritedness.

<sup>15</sup> Sumner, 1996, 146-147.

well against one's standards or expectations. The structure of Sumner's concept of being happy with one's life mirrors our *ex post* disposition to desire something and our taking ourselves to have reason to desire it. We could say along with Sumner that feeling enriched, rewarded, satisfied and fulfilled with one's life or some aspects of it is the *ex post* kind of affection that one could feel for one's life or some aspects of it, and that the disposition to have that affection is typically accompanied by the disposition to take oneself to have reason to have that affection. Sumner, however, takes this attitude to be a single one with both a cognitive and an affective component, while I would think that it is perfectly possible to desire something without thinking that one has reason to desire it and *vice versa*, one may take oneself to have reason to desire it something without desiring it.

#### 2.4. An example

Let us illustrate our position with an example. Suppose that it is an irreversible fact about my physical/mental constitution that I can never derive any pleasure from a certain kind of activity, say, listening to music. Suppose also that listening to music could never be subsumed under any other desirability features. No matter how much information I gather about music, I can never form a desire for listening to it, nor can I consider that activity to be good for me. I think that our account of well-being would have no difficulties in dealing with this kind of case. There is no reason for me to desire to listen to music; listening to music is not and cannot be part of my well-being. The reason-to-desire relation must obtain between an agent, and a set of facts (including facts about the agent), at a time. The fact that there is reason to desire pleasure does not in and of itself show that there is reason to desire any activity that is normally taken to be pleasant. Maybe a certain activity does not cause pleasure to arise in a certain agent. For this agent, the reason-to-desire relation cannot obtain.

Suppose now that the agent does not, at time  $t$ , derive any pleasure from listening to music but has a potential to do so, which would take some time to develop, and which he could have developed at  $t+1$ . Suppose that he has developed this potential. Does he have reason at  $t+1$  to desire listening to music at time  $t+1$ ? The answer is yes (saying that there is reason to desire it does not entail that there are no reasons of greater degree to desire other things at  $t+1$ ). Now go back to time  $t$ . Does the agent have reason to develop his potential? This would be an instrumental end towards the good of listening to music at  $t+1$ . The right

question to ask then should be: does the agent have reason at  $t$  to desire to listen to music at  $t+1$ ? The answer will depend on facts about the agent's future life. Given that the agent has the potential to enjoy music, and supposing that in his life he will have a chance to listen to music, and that listening to music would not otherwise be detrimental to him, then, I think we could suppose that there is reason for him to desire to listen to music and thus to develop his potential. If, on the other hand, it could be predicted that the agent, say, would never have a chance to listen to music, and in fact would derive great frustration from that fact, then it seems that there is no reason for him to desire to develop his potential. Finally: is there reason to desire at  $t$  to listen to music at  $t$ ? The answer is negative. We postulated that in order for music to be a good for the agent, developing a potential was required. Listening to music would otherwise yield no pleasure and, as stipulated, pleasure is the only desirability feature under which listening to music is subsumable.

### 3. The semantic role of desires

#### 3.1. Subjective or objective theories

Since Parfit's *Reasons and Persons*, scholars working in this area of philosophy lay a lot of stress on the distinction between subjective and objective theories of well-being. As with many distinctions in philosophy, however, there is no clear agreement on what the distinction is. As we shall see the question revolves around the role played by a person's desires and/or positive endorsements in determining the notion of well-being. Let us begin with a recent attempt to define objectivity. Arneson writes that a theory of well-being is *objective* just if it holds

that claims about what is good can be correct or incorrect and that the correctness of a claim about a person's good is determined independently of that person's volition, attitudes, and opinions.<sup>16</sup>

A theory of well-being is *subjective* if it denies just that. Two elements can be clearly discerned in Arneson's definition. (1) Claims about a person's good can be correct or

incorrect; and (2) the correctness of claims about a person's good is or is not determined by that person's attitudes towards the putative good in question. I take Arneson's definition to imply that both (1) and (2) must be true in order for a theory to be considered as objective. Now take a theory such as the desire-satisfaction theory of well-being:  $x$  is good for  $y$  iff  $x$  is the satisfaction of one of  $y$ 's desires. According to this theory, claims about well-being can be correct or incorrect, even though their correctness is entirely determined by the agents' positive attitudes (desires etc.). In this sense, on this theory claims about well-being could be correct or incorrect and some conditions for their assertibility are stated.

Desire satisfaction theories, then, would satisfy (1) but not (2). This second dimension is the real point of contention between philosophers who purport to defend 'objective' theories of well-being and philosophers who purport to defend 'subjective' theories. More specifically, the point of contention between subjective and objective theories focuses on what requirement or conditions should obtain for propositions about well-being to be correct. Broadly speaking, objective theories claim that the pro-attitudes of the agent towards an object are not necessary in order for that object to determine her well-being. On the other hand, for a theory to count as subjective the following must be true:

Nothing can intrinsically enhance the quality of a person's life unless that person has some positive, affirmative attitude toward that element of her life.<sup>17</sup>

Arneson refers to this constraint on theories of well-being as the weak endorsement constraint. This requirement has been taken very seriously. As Arneson puts it, what lends plausibility to this constraint is that a purportedly happy occurrence in one's life that leaves one utterly cold cannot intrinsically enhance one's well-being.<sup>18</sup>

---

<sup>16</sup> Arneson, 1999, 115.

<sup>17</sup> Arneson, 1999, 139. Note that this formulation of subjectivity is weaker than the one expressed in the previous quotation from Arneson. On that formulation it seemed that *subjectivity* entailed that a person's volitions and attitudes be both necessary and sufficient. On this weaker reading of subjectivity a person's volitions and attitudes are considered to be only necessary. Henceforth, we will refer to subjectivity in this weaker sense.

<sup>18</sup> To avoid this problem some hypothetical-desire theories place a constraint that any non-instrumental, basic desire whose satisfaction contributes to an individual's well-being must be actual when satisfied. That is to say, the hypothetical desire must be endorsed by the person and so become her desire in order for its satisfaction to contribute to her well-being. Griffin, 1986, 11.



On what side of this dichotomy does the reason-to-desire account stand? The short answer is that it stands on the objective side: there may be reason for an agent to desire something even though the agent does not desire or endorse that thing. Let me justify this position with an example. Suppose that a third person,  $y$ , correctly concludes that there is reason for  $x$  to desire  $g$ .  $x$ , however, has no desire for  $g$ . There are three different cases to be considered. (a)  $x$  may be unaware that he has reason to desire  $g$ , as he has failed to consider the issue at all; (b)  $x$  may have considered the issue but arrives at the incorrect conclusion that there is no reason for him to desire  $g$  (whether or not his rationality could be faulted for the wrongness of his conclusion); or (c)  $x$  can see that he has reason to desire  $g$  but that conclusion leaves him incapable of desiring it. I do not take it that there must be an *actual* desire for an object before the agent can enjoy this object at all. Take any of the three cases (a), (b) or (c). Suppose, for example that at  $t$ ,  $x$  has no (*ex ante*) desire for  $g$  and at  $t+1$   $g$  falls on  $x$ 's lap like manna fallen from heaven (maybe  $y$  secretly smuggles some  $g$  into  $x$ 's life). It might very well be that at  $t+1$ ,  $x$  rightly welcomes or responds positively to  $g$  even though at  $t$  he had no desire for  $g$ . The fact that  $x$  did not desire  $g$  at  $t$  does not make any difference to the fact that there was reason for him to desire  $g$  at  $t$  and there is reason to respond positively to the enjoyment of  $g$  at  $t+1$ .

It might be claimed, then, that what was important is not that there is an actual desire for  $g$  at  $t$ , but that once  $g$  is obtained it would actually be positively endorsed. What would this endorsement be? Arneson talks vaguely in terms of a "positive, affirmative attitude" towards an element of  $x$ 's life. I think one might be more precise and say that the endorsement in question is nothing other than the idea of the *ex post* desire discussed above. Suppose once more that  $g$  falls on  $x$ 's lap like manna from heaven. We can now ask: is this attitude of endorsement of  $g$  necessary in order to consider  $g$  as a good? The answer is once again going to be negative. As we saw above, the endorsement in question is a defeasible criterion to determine whether one has reason to desire something. Suppose the following three cases analogous to (a), (b), and (c) in the paragraph above. (a)  $x$  might not be aware that  $g$  is present in his life and cannot therefore endorse it; (b)  $x$  might be aware of  $g$  but conclude wrongly (and whether or not his rationality is faulty) that there is no reason to endorse it; or (c)  $x$  might see that there is reason to endorse it and yet fail to be emotionally engaged by it.

In conclusion, then, our theory would not accept the “weak endorsement constraint”. *Ex ante* and *ex post* desires are not necessary determinants of well-being. If that is true, however, on our theory it would be possible for an agent to attain something that intrinsically enhances his well-being though the agent does not have *ex ante* or *ex post* desires for it. It is not impossible to think about such cases in practice. As I shall later argue, a person’s well-being can be contributed to posthumously. Suppose, then, that Mike was a big Liverpool fan and until the tragic incident in which he lost his life he was very active in promoting the interests of his football club. After his death and despite the lack of a will, Mike’s relatives decide to donate on Mike’s own behalf part of his money for a new project in favour of the club. The relatives think that had Mike been aware of their decision he would have been very happy with it. They think that Mike’s interest is somehow best served in this way. Those of us who are inclined to believe that Mike’s well-being is to at least some extent being contributed to, may argue that that is so despite the lack of both an *ex ante* and *ex post* desire. It may be retorted that Mike’s dispositions (whilst alive) could clearly be understood as the expression of an *ex ante* desire for any project of the kind in question. Even if that were true, however, I would see no argument to the effect that these dispositions should be seen as *determinants* of Mike’s well-being rather than as an *indicator* of (that on the basis of which Mike’s relatives would *know*) what Mike would have had reason to desire.

On our reason-to-desire theory it would also be at least logically possible for an agent to get something that intrinsically enhances his well-being though the agent actually hates. I have to admit that I find it hard to come up with a good example to illustrate this case. There is, however, a good explanation for that, in terms of our own account: though desires are not determinants of well-being, they are our *primitive* criterion for claims about well-being. Without them, it will be hard to *know* anything about what is or what isn’t a contribution to a person’s well-being.

Note also, that the logical possibility in question is the same kind of logical possibility that explains how it would be possible that I don’t believe that *p* even when it is the case that *p*. The same analogous cases (a), (b), and (c) apply to the epistemological case too. I might for example not be aware of the evidence in favour of *p* and fail to believe that *p*. Or I might be aware of the evidence but incorrectly judge that it does not support *p*. Or I might be cognitively irrational and fail to believe that *p*, though I realise that the evidence

supports *p*. We would take it to be a rather worrying sign if a theory of well-being did not allow for even the logical possibility of mistakenly hating something that, in fact, is good for the person.

### 3.2. Does the reason-to-desire account leave no room for individual differences?

From this discussion, we must conclude that our account is on the side of objective theories of well-being, at least insofar as one agrees with Arneson's distinction between objective and subjective theories. Let us make clear, however, that this type of objectivity should not trigger liberal fears of paternalism. True enough, on our theory there can be only so many desirability features under which there can be reason for any agent to desire an object or a state of affairs, and criteria of what these features are, are provided by looking at the converging judgements of authoritative judges. We have mentioned pleasure, so far, and believe *contra* Mill that that is not the only feature that competent judges would single out as desirable for one. Experience and reflection, I believe, will show that knowledge, autonomy, deep personal relations, achievement, and virtue can also be added to pleasure. However, our theory does not determine a list of *particular* objects and states of affairs that are to be desirable for each human agent. To say that there are only so many desirability features is to say that each *particular* object or state of affairs an agent has reason to desire is to be subsumable under a determinate *categorical* end, be it pleasure, knowledge, or any other categorical end on which the opinion of competent judges would converge.

Categorical ends must be intelligible. If we observed a person eating dirt with a spoon, we would want to understand his action and in order to do so we would begin by trying to place it under any one of the categorical ends that it is intelligible to apply to human actions. Does this person find eating dirt pleasant? Or is he rather trying to acquire some knowledge, maybe, the knowledge of what dirt tastes like? If he isn't what makes his action desirable or valuable for him? In order to be able to understand these actions we would have to place them under kinds of ends that we spontaneously find intelligible, ends that we as human beings can share. If we can't place these actions under any intelligible end, we cannot but deem them irrational.<sup>19</sup>

Saying that there is no objective list of *particular* objects and states of affairs that each and every agent should find desirable is compatible with the claim that given some general facts about human biology, certain general claims about the good for human beings can be made with a certain degree of confidence. This claim, however, would not amount to claiming that, say, there is more reason for each and every human being to desire chocolate ice cream over vanilla ice cream, or to desire listening to jazz music over pop music. Whether there is reason for a certain agent to desire one over the other will depend on some facts about the agent in question. In this case, supposing that these two kinds of ends are entirely subsumable under the idea of pleasure, what there is more reason for the agent to desire will depend on what option is likely to yield more pleasure. At this level, it will be some fundamental fact about the agent that determines what there is reason for him to desire. What is more, at this level, the agent is likely to be the most authoritative judge to determine which of the two options is best for him.

Notice also that on my account this sensitivity to individual differences might be reproduced at the more fundamental level of the categorial ends. So far, I have only claimed that there are a number of *intelligible* categorial ends. What I have not specified is (1) whether there is a ranking of these ends in order of prudential importance applying to each individual; and (2) what consequences the presence of such ranking would have on what there is reason for each agent to desire.

Let us begin with question (1). It might not be implausible to postulate that certain categorial ends rank higher than others. This conclusion might be more easily accepted by those who take it that there may be higher and lower kinds of pleasure. The argument then would use the following analogy: just as there are higher and lower pleasures, there are higher and lower categorial ends. Of course, this argument is nowhere near to being a very compelling argument. For one thing, the term of comparison on which the analogy is supposed to rest (that there are higher and lower pleasures) is very much in dispute. Supposing, however, that it was unanimously agreed that certain pleasures are higher than others: what would that exactly entail for each individual's well-being? Suppose that a competent judge were to rank two goods,  $g$  and  $g^*$ . The judge would have to be well acquainted with both goods; the spontaneity of her desires for, respectively,  $g$  and  $g^*$

---

<sup>19</sup> Of course, we might refrain from drawing this conclusion if the actions we found unintelligible were those

should be genuine and not to be faulted; and finally, her judgement should converge with that of the majority of competent judges. Suppose that she correctly judged  $g^*$  to be a superior kind of good. Would that entail that there is reason for each and every agent to desire  $g^*$  over  $g$ ? Certainly not. Suppose, for example, that it is a fact about agent  $x$  that he has no access to  $g^*$ . Maybe  $g^*$  is very expensive or maybe  $x$  has no access to  $g^*$  because he lacks the physical/mental capacities to appreciate  $g^*$ . In either case, given these facts about  $x$  it would certainly be wrong to assert that there is reason for  $x$  to desire  $g^*$  over  $g$ .<sup>20</sup> The same point applies *pari passu* to the case of categorial ends. Even if they could be ranked in order of prudential importance, this would not entail that each and every agent, irrespective of his or her situation, would have more reason to desire the objects and states of affairs subsumable under the higher categorial ends.

In conclusion, the reason-to-desire theory of well-being as specified here holds that judgements about well-being can be correct and incorrect and that desires or endorsements are not necessary determinants of an agent's well-being. This, however, does not entail that there is equal reason for each agent to desire the same particular objects or state of affairs. What an individual has reason to desire always depends on facts relative to him or her, though of course human beings share many (psychological, physiological) features *qua* human beings and it is thus possible to make some generalisations with a good degree of confidence. We have shown that our account leaves comfortable room for differences in taste and for the fact that at least when it comes to taste, each individual might be the best judge of what is good for him or her. Also, it should not be forgotten that on this account the most important epistemic criteria of desirableness are each person's desires and endorsements. In most cases, agents will not be 'left cold' by something that is good for them. If they do after having duly pondered the facts, they might be cognitively or affectively criticisable.

#### 4. Ideal-preference v. reason-to-desire

---

of beings very unlike human beings.

<sup>20</sup> Though one could say that there *would be* reason for her to desire it *if* some relevant facts about her and her circumstances were different.

The reason-to-desire theory of well-being might seem very close to what we could generally refer to as ideal-preference theories of well-being. It will be the task of this section to shed light on what I take to be the unbridgeable distinction between them. We shall begin by briefly presenting the development of ideal preference and reason-to-desire theories of well-being as they were discussed by Sidgwick. Subsequently I will discuss some of the difficulties that arise for ideal preference theories, which, as I shall show, do not arise for the reason-to-desire theory of well-being.

#### 4.1. Ideal preference theories

One of the earliest discussions of both kinds of theories in the modern debate can be found in Sidgwick's *Methods of Ethics*. It is important to place Sidgwick's discussion in its context. Sidgwick believed that practical judgements such as judgements about the good are considered by common sense to be determined by rational considerations or dictates. He also assumed that the grammatical forms of ordinary language can tell us something important about the nature of what we talk about. In his discussion about this type of practical judgement, Sidgwick examines the contrary opinion of some philosophers according to whom such judgements are either expressions of desires or inclinations or else factual judgements containing information to which our desires lead us to react favourably or unfavourably.<sup>21</sup>

Sidgwick begins by exploring and rejecting the idea that the notion 'good' can be interpreted in relation to simple desires. He argues that that will not work for two reasons: firstly because "what is desired at any time is, as such, merely apparent Good, which may not be found good when fruition comes, or at any rate not so good as it appeared."<sup>22</sup> Secondly, he argues that prudent men would diminish the intensity of their desires for things commonly considered as being good for us such as fine weather or perfect health. That is because these things lay outside our reach. To desire them intensely would generate great frustration, i.e., something not good for one. Hence, prudent men attempt to diminish the intensity of their desires for these things; that, however, does not make these things any less good.<sup>23</sup> Thus, Sidgwick considers whether we should:

---

<sup>21</sup> For a detailed discussion of this topic see Schneewind, 1977, 215-226.

<sup>22</sup> Sidgwick, 1907, 110.

<sup>23</sup> Sidgwick, 1907, 110.

identify [the notion of 'good'] not with the actually *desired*, but rather with the *desirable*;—meaning by 'desirable' not necessarily 'what *ought* to be desired' but what would be desired, with strength proportioned to the degree of desirability, if it were judged attainable by voluntary action, supposing the desirer to possess a perfect forecast, emotional as well as intellectual, of the state of attainment or fruition.<sup>24</sup>

But Sidgwick finds this definition unsatisfactory, as it limits the scope of our judgements of good to what is within our reach and it overlooks the fact that there are things that we desire and that we do not regret having obtained, but which destroy certain desires which could have brought about something better for the agent. Put simply, this definition cannot account for the 'good on the whole' of an agent. So, on behalf of those who want to maintain that practical judgements are either expressions of desires or inclinations or else factual judgements containing information to which our desires lead us to react favourably or unfavourably, Sidgwick put forward the following thesis, which we will call (S):

- (S) A man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point in time.<sup>25</sup>

And he commented on it as follows:

The notion of 'Good' thus attained has an ideal element: it is something that is not always actually desired and aimed at by human beings: but the ideal element is entirely interpretable in terms of *fact*, actual or hypothetical, and does not introduce any judgement of value, fundamentally distinct from judgements relating to existence;—still less any 'dictate of Reason'.<sup>26</sup>

At this point, Sidgwick begins to criticise (S) though, as we shall see, his criticism is not as clear cut as one would expect it to be. He first tells us that though (S) is clearly not "what we commonly *mean* when we talk of a man's 'good on the whole'" it cannot be denied

---

<sup>24</sup> Sidgwick, 1907, 110-111.

<sup>25</sup> Sidgwick, 1907, 111-112.

<sup>26</sup> Sidgwick, 1907, 112. Of course, the "actual" in this quotation does not refer to (S) but to Sidgwick's earlier discussion of definition of the good in terms of actual desires. (S) is clearly expressed in terms of hypothetical facts.

“that this hypothetical object of a resultant desire supplies an intelligible and admissible interpretation of the terms ‘good’ (substantive) and ‘desirable’, as giving philosophical precision to the vaguer meaning with which they are used in ordinary discourse.”<sup>27</sup> However, it still seems to him “more in accordance with common sense to recognise—as Butler does—that the calm desire for ‘my good on the whole’ is *authoritative*; and therefore carries with it implicitly a rational dictate to aim at this end, if in any case a conflicting desire urges the will in an opposite direction.”<sup>28</sup> As we saw, Sidgwick did not consider (S) to express this rational dictate.<sup>29</sup> To do just that, Sidgwick put forward what we shall call (S\*):

(S\*) ... ‘ultimate good on the whole for me’ [means] what I should practically desire if my desires were in harmony with reason, assuming my existence alone to be considered.<sup>30</sup>

According to Sidgwick, then, the relevant difference between (S) and (S\*) is precisely that the latter alludes to the *authoritative* dictates of reason implicit in our common judgements about the good, while the former doesn’t. This point should be read in conjunction with Sidgwick’s earlier remark to the effect that (S) expresses judgements about the good in terms of judgements “related to existence”, i.e., descriptive judgements while, as we are to gather, (S\*) does not imply that. Read in this way, then, Sidgwick seems to be saying that the normativity of judgments about the good cannot be captured by descriptive claims about our desires even when these are formed under idealised conditions. Hence, Sidgwick’s remark about conflicting desires: when I judge that something is good for me, I recognise that I should desire that thing whatever else is true about my desires actual or hypothetical. These actual or hypothetical desires may simply fail to carry any normative weight. However, though at a *semantic level* judgements about well-being might not be expressed in terms of what one would desire under idealised conditions, at an *epistemological level*, it might still be the case that the best way to find out what is good for one is to think about what one would desire under idealised conditions.<sup>31</sup> Sidgwick’s account is not inconsistent with this last point.

---

<sup>27</sup> Sidgwick, 1907, 112.

<sup>28</sup> Sidgwick, 1907, 112.

<sup>29</sup> Christiano, 1992, also noticed Sidgwick’s dissatisfaction with (S)’s fitness to express the “rational dictate”.

<sup>30</sup> Sidgwick, 1907, 112

<sup>31</sup> I am not claiming, however, that Sidgwick himself endorsed this evidential view. Christiano writes that Sidgwick never talked of desires as evidential, though he thinks that Sidgwick’s view is best made sense of when understood in these terms. Christiano, 1992, 264.



At this point two remarks ought to be made. The first one is only peripheral to our aim and exegetical in nature. I believe our presentation of Sidgwick's thought to be faithful both to the relevant parts of his text in particular, and to his methods and projects in general. As for the latter, I am particularly referring to his intuitionistic defence of the importance of ordinary language in pointing to the nature of the things we talk about. It seems clear that Sidgwick, though somehow attracted by (S) as a fine philosophical interpretation of the notion of the good for one, thought that (S) was not capable of fully capturing the meaning of our common notion of the good. His formulation of (S\*) ought to be read precisely in this light, as a remedy to the shortcomings of (S).<sup>32</sup>

Secondly, it has to be noted that the recent literature on the ideal preference account offers different versions of such an account. The common idea behind them all is that well-being is to be characterised by starting with a desire (or with a preference or with contentment) that is idealized by attaching some condition C, be it avoidance of the influence of mistaken belief and of ignorance of material facts, etc.<sup>33</sup> Condition C may be spelled out in such a way as to produce a *reduction* of the normative concept of well-being to a non-normative concept, i.e., a desire or preference of some kind. We have seen such an attempt with Sidgwick's (S) above. However, condition C may also be spelled out in a way which Sidgwick did not consider, i.e., so as to produce an a priori equivalence between two normative concepts. One example would be the following:

*g* is an ingredient of *x*'s well-being if and only if *g* is what *x* would desire if *x*'s rationality could not be faulted and *x* was equipped with all relevant information.

---

<sup>32</sup> I should at this point note that my reading is by no means shared by all of the students of Sidgwick's thought. Many authors fail to have noticed the relevant difference between (S) and (S\*). Brandt, 1979, 59, 155; Darwall, 2002b, 31-38; Griffin, 1986, 39; Sobel, 1994, 792, 795, n. 23; Rawls, 1971, 416-17, 421; Rosati, 1995, 297-98. Shaver, 1997, 314 n. 2, also mentions Loeb, 1995, 24, n. 2; and Velleman, 1988, 353-54. Some authors, however, have noticed the difference in question: Parfit, 1984, 500. Schneewind, 1977, 225; Shaver, 1997, 314-17. Finally, Crisp, 1990, 270, notices the difference between what I formulated as (S) and (S\*) but claims these are not two separate alternatives but collateral definitions of the Good: (S) "spells out what it is that makes someone's life go best" while (S\*) "implies that seeking this is what one has most reason to do." What I claim, however, is that Sidgwick noticed that what there is most reason to desire (and hence do) might not be what one would desire under idealised conditions. If that's correct, Sidgwick should not have taken (S) to spell out what it is that makes someone's life go best.

<sup>33</sup> Griffin, 1986, chs. 1-2; Brandt, 1979, chs. 6-7; Rawls, 1971, 417 ff. Brandt for example defines as rational the intrinsic desires that would survive the idealised scrutiny that he describes under the notion of 'cognitive psychotherapy': an intrinsic desire is rational if a person would still have it after repeated representation of all relevant scientifically available information, in an ideally vivid way, at appropriate times.

Even this formulation, however, would preserve the difference between ideal preference theories and the reason-to-desire theory. On our account, though these idealised desires play a very important role at the epistemic level, we shall say (in Sidgwickian spirit) that they can never afford a characterisation of well-being at the semantic level. This would of course be precisely denied by the defenders of ideal preference theories of well-being. Let me dwell on this point a little further.

The reason-to-desire relation is a relation that obtains when for example the fact that  $p$  gives  $x$  reason (of degree  $d$  at time  $t$ ) to desire  $\phi$ . Suppose that the fact that smoking gives me pleasure is a reason for me to desire to smoke. That there is reason for me to desire to smoke is true in virtue of the fact that there is reason to desire pleasure, and some facts about me, namely, that I am such that smoking would give me pleasure. The spelling out of the relata indicates the normative statement that smoking is desirable for me. In the sections above, we described the epistemology of these purely normative statements: a spontaneous judgement about reasons stemming from the feelings must be rationally validated in order to become a *warranted* normative judgement. The process described above occurring from our spontaneous feelings through to their validation as warranted normative judgements did not offer a *definition* (strict analysis) or a characterisation (a priori equivalence) of what there is reason to desire in terms of a counterfactual. It is precisely on this point that the reason-to-desire account and the ideal-preference account differ most clearly.

#### 4.2. Problems for ideal preference theories

Robert Shope formulates a serious objection against definitions of rational desires expressed in terms of subjunctive conditionals like the ideal preferences we have been discussing so far.<sup>34</sup> His point is best illustrated through the following example. Suppose that I strongly desire to learn ethical theory but if I were fully informed about all the relevant matters concerning the reasonableness of my choice, I would already know ethical theory and hence would have no desire to acquire further understanding of it. Learning about ethical theory, then, would not be part of my well-being as I would not form such

---

<sup>34</sup> Shope, 1978a; 1978b.

desires under ideal conditions, and this is indeed a counterintuitive way to show that something is not in my interest. Even if learning about ethical theory were not part of my well-being it is surely not so for the reason ideal preference theories seem to suggest.

Donald Hubin makes a similar remark concerning counterfactual accounts.<sup>35</sup> Ideal preference accounts, or as he would call them, hypothetical motivation accounts of well-being, claim that we have reason to desire to do that which we would be moved to desire to do under certain conditions. Our reasons to choose, desire, prefer, be content with etc., an object or a state of affairs *is* a disposition to choose, desire, prefer, be content with etc. that object or state of affairs under certain conditions. It is this disposition that the subjunctive conditional defining the ideal or hypothetical preference is supposed to capture. But, then, the following problem arises:

Let us suppose that the idealizing conditions the theory proposes are these: removal of false beliefs, introduction of all relevant true beliefs, and vivid appreciation of all relevant factors. Now, we are to ask what a real flesh-and-blood agent *would* be motivated to do [i.e., choose, desire, prefer etc., to do] were he to be free of false beliefs, vividly aware, ... *etc.* But what a human being would be motivated to do under non-actual circumstances ... is in part dependent on brute physiological features of the agent. This particular agent, even when idealized in the ways we are imagining, may become motivated to put pebbles in his navel when he becomes vividly aware of some complex of facts about number theory. The awareness of the set of facts may produce physiological effects on the agent—effects that have nothing to do with the content of the facts—and these effects may produce other psychological states: desires, or some other psychological states taken to be motivating.<sup>36</sup>

In this case, the hypothetical account would suggest that the agent has a reason to desire to do something that there is no reason to desire to do. Hubin considers and rejects various ways in which the defenders of the hypothetical account could reply. For one, the idealizing conditions might not have been the right ones. Hubin convincingly shows, however, that his example applies across the board to different types of idealisations.<sup>37</sup> Alternatively, it might be replied that those counterfactuals based on purely *physiological*

---

<sup>35</sup> Hubin, 1996. In what follows I will not strictly reproduce the terminology used by Hubin's, as it is different from ours. His points can be expressed in our own terminology with no loss of meaning.

<sup>36</sup> Hubin, 1996, 36-37.

<sup>37</sup> See Hubin, 1996, 37.

causal links as that in the example above should be excluded. Hubin then goes on to consider the case in which the hypothetical circumstances focus only on the *psychological* (rather than physiological) connections based on the contents of mental states. He concludes that even when considering *nondeviant* (normal) psychologies, people might be such that vivid awareness of certain facts would trigger a desire (choice, preference, etc.) to do something that they clearly have no reason to desire. Vivid awareness of the horrors of the Holocaust, for example, would kill some people's desire for food. Hubin thus concludes:

motivation [or preference, choice, desire, etc.] might withstand (or be extinguished by) the idealization for the wrong reasons; the idealizing conditions may even create inappropriate motivation.<sup>38</sup>

The reason-to-desire account does not fall into this kind of problem because what there is reason to desire is not *defined* (nor is it an a priori equivalence) in terms of what the agent *would* desire under certain conditions. Ideal preference theorists might reply to Hubin's charges by laying more stress on the agent's rationality under the idealised conditions so that it could not be possible for the agent to have an ideal preference for anything other than that which he has reason to desire. It may for example be argued that the agent is rational only when he recognises all the truths about the reason-to-desire relation. But this would make the ideal preference account redundant. It would be simpler to appeal directly to the reason-to-desire relation.

---

<sup>38</sup> Hubin, 1996, 44. Before Hubin, Gibbard makes a similar point when he writes: "Perhaps with a more vivid realization of what peoples' innards are like, I would want to stay away from dinner parties and do all my eating alone—although then I would feel lonely and isolated." Gibbard, 1990, 20. Gibbard concludes that full-information theory of rationality would define as rational the desire to eat alone, and yet this is not what would be advisable for one to do. As a consequence, rationality would lose its advisability or recommending force. This last point links Gibbard's discussion to Sidgwick's claim to the effect that (S) cannot fully capture the 'Rational dictate' implicit in the notion of 'the good on the whole' for a person.

In order to avoid this problem Peter Railton suggests the following formulation: "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality." Railton, 1986, 16. This account is also known as the ideal advisor account of the good. I believe Railton's account cannot successfully deal with the Gibbard-Hubin point. Suppose that I am motivated to put pebbles in my navel when I consider under idealised conditions whether it would be good for me to write a Ph.D. in ethical theory. My ideal advisor would know this fact about me and discount it. But on the basis of what can he discount it? Not on the basis of the idealised conditions as those conditions are precisely what yield the fact to be discounted.

## 5. Conclusion

In conclusion, on the reason-to-desire theory, what measure of well-being each agent attains depends on the extent to which his life contains what he has reason to desire. The measure of well-being I enjoy increases the greater the measure of pleasure, achievement, knowledge, deep personal relations, autonomy I enjoy. The distribution of these goods that will produce the greatest measure of well-being will vary for each agent according to facts about his or her person and his or her particular circumstances. Judgements about what there is reason to desire can be authoritative: *ex ante* and *ex post* desires are the primitive criteria for these judgements. These criteria are typically accompanied by a normative impulse: we take ourselves to have reason to desire the objects of our desires. Reflection, experience and the converging opinion of competent judges confer authoritativeness on our initial judgement. This account need not mention the satisfaction of desires, and preferences, whether ideal or not. In the next three chapters, the plausibility of our theory of well-being will be put to the test.

## Chapter 5                      The Scope Problem and the Limits of Well-Being

### 1. Introduction

In this chapter I examine issues concerning the limits of the concept of well-being. I begin by looking at Stephen Darwall's rational care account of well-being. According to Darwall an account of well-being is too wide when it includes as contributions to a person's well-being instances of rational self-sacrifice. If it does, Darwall claims there is a "scope problem": we must find an account on which well-being has narrower limits. Though I agree with Darwall's aim, I disagree with his strategy at least when it consists in claiming that well-being is not a normative concept and that the notion of care is what lends normativity to well-being (§2). Next, I illustrate how our reason-to-desire account deals with the scope problem (§3). This discussion will yield the conclusion that a person's well-being can be contributed to by things as diverse as the well-being of other persons, all sorts of projects the outcomes of which (in one sense) do not involve one's self, and even post-mortem events and states of affairs. It might be wondered, then, whether we could still call 'well-being' something as wide in scope as this concept (§4). Finally, I will deal with a difficulty raised by our claim that a person's well-being can be contributed to by post-mortem states and events (§5).

### 2. Darwall, the scope problem, and the normativity of well-being

#### 2.1. The scope problem

Darwall's particular view of well-being seems to be motivated by a worry which I share. Darwall wants to find a theory of well-being capable of accommodating the possibility of rational self-sacrifice. The idea intuitively at work in his account is the following: when a person makes a rational self-sacrifice he cannot at the same time be doing what is best in terms of his well-being and *vice versa*. As I would put it, the ideas of rational self-sacrifice

and well-being are essentially connected: an action cannot rationally be conceived of as both what's best for  $x$  in terms of his well-being and a rational self-sacrifice for  $x$ .<sup>1</sup>

According to Darwall, we must distinguish between on the one hand a theory of rational choice and, on the other, a theory of well-being. Failing to do this opens up the door to what Darwall calls the *scope problem*: if rational choice theories are conflated with rational self-interest theories the scope of the concept of well-being would be so wide as not to allow the possibility of rational self-sacrifice. Darwall thinks that informed desire accounts of the good might very well work as wider accounts of rational choice. The fact, however, that on this type of account it might be rational for an agent to sacrifice his well-being on one occasion, shows that they cannot be considered as accounts of individual well-being. We must draw a distinction between rational choice and rational self-interest the dividing line being the notion of rational self-sacrifice.

Even though I am in agreement with Darwall's aim, i.e., giving an account of well-being narrow enough in scope so as to exclude that instances of rational self-sacrifice may count as part of one's well-being, I am opposed to the way in which he argues for it. Darwall's strategy can be inferred from the following passage:

But what is a person's good, welfare, or interest? And what claims does our own good make on us? These questions are interrelated. So long as the concept of a person's welfare is seen as having a normativity that is intrinsically agent-relative, as entailing normative reasons distinctively for the *agent's own* desires and actions, it will seem inevitable that welfare consists in whatever it makes sense for a person to want and, therefore, to realize. To a first approximation, what is *in* a person's interest will seem to consist in whatever an agent (rationally) *takes* an interest in. This, however, make self-sacrificial acts... impossible.<sup>2</sup>

Darwall's strategy, then, consists in showing that the concept of well-being is not in itself an agent-relative normative concept. If well-being can be shown not to be intrinsically normative, then saying that something is good for a person does not amount to saying that

---

<sup>1</sup> We shall discuss the notion of self-sacrifice in greater detail in Chapter 7.

<sup>2</sup> Darwall, 2002b, 23-4. Note that Darwall seems to believe that if it makes sense to desire something, it make sense to realise the object of one's desire. Though I agree with this thesis, I do not take it to imply that if it

he or she has reason to desire and pursue it or “take an interest in it”. As we shall soon see, Darwall wants to show that the notion of care is what confers on well-being its normativity.

## 2.2. Presenting Darwall's challenge: care and the normativity of well-being

Darwall writes:

...it is believed that if something is for my own good, then it follows that I ought, or have reason, to want or pursue it. It is assumed, that is, that welfare has an *agent-relative* normativity, that a person's welfare is necessarily normative for *his own* desires and actions.<sup>3</sup>

Let us express his thought as the following thesis:

(W) (*y*) (If something is for *y*'s good or well-being, then *y* has reason to want it and pursue it)

Darwall intends to deny that (W) is a conceptual truth. The bulk of his argument is expressed in the following two passages:

...someone may not value his own welfare because, in a depression, he sees himself as *unworthy* of care or even, perhaps, because he loathes himself. Depression or self-loathing doesn't entirely extinguish values and preferences, however. The depressive may prefer isolation and sleep, even though he knows that he might enjoy and benefit more from going out with friends: “Sure, that would be better for *me*,” he might say “ but why does that matter? Why think I am worth caring for?” And the self-loather might take the fact that he

---

makes sense to realise something, it makes sense to desire that thing. This point should be kept firmly in mind for the rest of this discussion.

<sup>3</sup> Darwall, 2002b, 4. Two connected remarks concerning this quotation and its formalisation are in order. Firstly, note that according to Darwall it is believed that (a) if something is good for someone than he has reason to want *or* pursue it and (b) a person's welfare is necessarily normative for his own desires *and* actions. It is clear from the discussion following this quotation that Darwall takes it that we have reason to *pursue* something when we have reason to *desire* it and *vice versa*. Of course, and this is my second remark, this is not what I have claimed: though we have reason to pursue or realise the things we have reason to desire, it is not the case that we have reason to desire to do all the things we have reason to do.



would benefit from an activity as a reason *not* to engage in it. To both, the thesis that one's own good or welfare entails reasons for acting will mock the truth.<sup>4</sup>

Most would agree, of course, that the depressive and the self-loather are mistaken in thinking that considerations of their own welfare give them no reasons. But what these characters think isn't self-contradictory or conceptually incoherent. And because it isn't, the normativity of welfare cannot consist in entailing agent-relative reasons for the person's own desires and actions. The notion that, as one is unworthy, one's good gives one no reasons, is not the incoherent thought that what is (as one thinks) valuable, gives one no reasons. It is *conceptually* possible that considerations of one's own good provide no normative reasons for acting whatsoever or even, as the self-loather believes, that they provide "counter" reasons. To claim otherwise, as I assume we would, is to put forward a substantive normative thesis, not an analytic or conceptual truth.<sup>5</sup>

Darwall thinks that (W) is not a *conceptual* truth, but wants to assert that another thesis is conceptually true:

To understand the normativity of welfare, I shall argue, we must see it in relation to care. What the depressive is right about is that *if* he weren't worth caring for, considerations of his own good would not be reasons. It's just that he is wrong in thinking he is unworthy of care. The deep truth that underlies the depressive's claim is that it is a person's being worthy of concern (as he will seem to someone who actually cares for him) that makes considerations of his welfare into reasons. What *is* a conceptual truth is that to care for someone is to be in a relation to him such that considerations of that person's welfare are normative for one's desires and actions with respect to him. *What is for someone's good or welfare is what one ought to desire and promote insofar as one cares for him.*<sup>6</sup>

Exactly what thesis is Darwall asserting to be conceptually true? The last quotation reports a paragraph from beginning to end. At the beginning of the paragraph, and in light of the preceding discussion, Darwall is clearly in the process of characterising the *normativity of well-being*: we can understand the normativity of well-being only when we "see it in relation to care". Darwall proceeds by specifying the relation between care on the one side

---

<sup>4</sup> Darwall, 2002b, 5-6.

<sup>5</sup> Darwall, 2002b, 6.

<sup>6</sup> Darwall, 2002b, 6-7.

and the normativity of well-being on the other. The penultimate sentence of the paragraph tells us that what *is* a conceptual truth is that:

- (d) “to care for someone is to be in a relation to him such that considerations of that person’s welfare are normative for one’s desires and actions with respect to him.”

The very last sentence of the paragraph, however, witnesses an unmentioned shift in the object being characterised: what is at issue now is *not* the normativity of well-being but rather well-being itself: “*What is for someone’s good or welfare is...*” Something Darwall writes later on may be taken to clarify this shift:

So far we have it that the normativity of welfare must be understood in relation to a concern for someone for that person’s sake. I will be claiming, in addition, that a stronger relation exists between welfare and care, namely, that what it is for something to be good for someone *just is* for it to be something that one should desire for him for his sake, that is insofar as one cares for him. ... We might equivalently say that what it is for something to be good for someone is for it to be something that is rational (makes sense, is warranted or justified) to desire for him insofar as one cares about him. This is a *rational care theory of welfare*. It says that being (part of) someone’s welfare is being something that it would be rational to want for him for his sake.<sup>7</sup>

It seems as if Darwall wants to defend two theses: the first is (d) and is concerned with the normativity of well-being and its relation to care; the second is meant to provide a characterisation of well-being itself. Let us state the latter thesis as follows:

- (d\*) “Being (part of) someone’s welfare is being something that it would be rational to want for him for his sake.”

In conclusion, Darwall makes at least three claims: (1) (W) is not conceptually true; (2) (d) is conceptually true; (3) (d\*) is (conceptually) true. In §2.3, I take a closer look at (d) and attempt to specify precisely which thesis Darwall is putting forward as a substitute for (W). I conclude that, at least on some interpretation of it, (d) yields some unintuitive results and

---

<sup>7</sup> Darwall, 2002b, 8-9. To fully understand this passage we must mention that according to Darwall “When we care about someone, we desire things for her for her sake.” Darwall, 2002b, 47.

wonder whether such a thesis can really be considered to be conceptually true. In §2.4, I will put forward reasons to believe that (d\*) is uninformative. Finally, in §2.5, *contra* Darwall, I argue that the best explanation of the depressive and the self-loather's case does not yield the conceptual possibility that well-being entails no reasons.

### 2.3. The normativity of well-being: which thesis?

Consider (d):

- (d) "to care for someone is to be in a relation to him such that considerations of that person's welfare are normative for one's desires and actions with respect to him."

The purpose of (d) is to show that the notion of care is what confers normativity to considerations of well-being. According to Darwall, well-being does not entail reasons unless we see it in relation to care. It is because  $x$  cares about  $y$  that the fact that something is good for  $y$  entails a reason for  $x$  to desire it for  $y$ . We could formulate (d) as follows:

- (D)  $(x), (y)$  (If something is for  $y$ 's good then, if  $x$  cares for  $y$ ,  $x$  has reason to desire and promote it)<sup>8</sup>

If I care about myself I will find that considerations of my well-being entail reasons for me. It is hard to argue against this train of thought except on grounds of circularity. We will pursue that line of argument in the next subsection. First, however, we must bring the reader's attention to another feature of Darwall's account.

Darwall might argue that (D) is not really the thesis he was trying to defend. In a passage quoted above, Darwall writes that "What the depressed is right about is that *if* he weren't

---

<sup>8</sup> Throughout this section I will take  $x$  and  $y$  to range over persons, i.e., agents with a capacity to think and act on reasons;  $x$  and  $y$  can be identical. It might be helpful to identify the logical form of (D). There are three atomic sentences "P", "Q", "R": respectively, "Something is for  $y$ 's good or interest", " $x$  has reason to desire and promote it", and " $x$  cares about  $y$ ". The logical form of (D) must be: " $P \diamond (R \diamond Q)$ ". Notice that this formulation is equivalent to " $R \diamond (P \diamond Q)$ ", i.e., if  $x$  cares about  $y$  then, if something is good for  $y$ ,  $x$  has reason to desire and promote it. This formulation may seem even closer to Darwall's claim that "to care for someone is to be in a relation to him such that considerations of that person's welfare are normative for one's desires and actions with respect to him."

worth caring for, consideration of his own good would not be reasons. It's just that he is wrong in thinking he is *unworthy of care*. ...[I]t is a person's being worthy of concern...that makes considerations of his welfare into reasons." This quote highlights an important shift in Darwall's conception of the normativity of well-being. On this new reading Darwall is not concerned with care as a crude (unevaluated) feeling: rather, what would make considerations of well-being normative is a person's being *worthy of care*. Then, rather than (D), the thesis at hand would have to be (D'):

(D') (x), (y) (If something is for y's good then, if y is worthy of care, x has reason to desire and promote it)

This is a very important change, as caring about y might have nothing to do with y's being worthy of care. Suppose that Jack the Ripper is someone completely unworthy of care: he is a serial killer, with no exculpating elements in his psychological profile. This fact does not exclude that someone (including himself) could care about him. But now let us look at the truth of (D'). Darwall would have to claim that it is a conceptual truth that the *agent-neutral* reason giving force of a person's well-being is conditional on considerations of worthiness of care. Suppose now that the building of a gigantic marble statue representing my person would immensely contribute to my good and, despite my eccentricities, I am someone who is worthy of care. Is this fact (my being worthy of care) really sufficient to make it the case that everyone has a reason to desire and promote the building of my statue. Rather than being conceptually true (D') seems substantively false.

#### 2.4. The definition of well-being in terms of care

Remember Darwall's definition of well-being:

(d\*) "Being (part of) someone's welfare is being something that it would be rational to want for him for his sake"

Darwall thinks that this formulation is equivalent to another one: "what it is for something to be good for someone is for it to be something that is rational (makes sense, is warranted or justified) to desire for him insofar as one cares about him." That is because "When we

care about someone, we desire things for her for her sake.”<sup>9</sup> Let us formulate (d\*) as follows:

(D\*) (x)(y)(g) (g is good for y iff, if x cares about y then it would be rational for x to desire g for y’s sake)

For example, suppose that Jack is identical to *x* and *y*, and *g* is the eating of an apple by Jack. Eating the apple is good for Jack, if and only if, if Jack cared about himself it would be rational for him to desire that he ate the apple. Note that, unlike the discussion concerning the normativity of well-being, the discussion concerning the definition of well-being is conducted by Darwall in terms of care rather than worthiness of care. In what follows, then, we shall ignore concerns arising from definitions of well-being in terms of worthiness of care. According to Darwall, we cannot define well-being simply in terms of what it would be rational to desire because the scope of ‘desire’ is too wide.<sup>10</sup> Care too is a form of desiring but with a very important qualification: when we care about someone we desire something for this person *for his sake*. As Darwall writes:

To appreciate what these last three words [*for his sake*] add consider that it seems possible for an intrinsic desire for someone to arise through the sort of associative process by which Mill explains the genesis of an intrinsic desire for wealth, or even, perhaps, through whim or fancy, without involving any concern for the person himself. ... One might simply desire intrinsically *that* another’s good be realized without desiring this for *his* sake.<sup>11</sup>

We might for example meet a stranger on a train and form an intrinsic desire that certain states concerning this stranger’s good obtain without desiring it for his sake. On the other hand:

The object of care is the person *herself*, not some state or property involving her. In caring for her, we, of course, want certain states and properties involving her to be realized. When they derive from care, however, such desires have an “indirect object” in addition to these direct objects. In caring for her, we want these things *for her*. This does not mean just that

<sup>9</sup> Darwall, 2002b, 47.

<sup>10</sup> According to Darwall it may for example be rational for someone to desire to self-sacrifice. See our discussion in §3 below.

<sup>11</sup> Darwall, 2002b, 2.

we want the properties and states to involve her, as opposed to someone else—or even that the desires are to be understood *de re* rather than *de dicto*. We want them for her *for her sake*.<sup>12</sup>

I take it that an intrinsic desire for something is a desire for that thing for its sake and not for the sake of something else. Suppose now that the only thing that would contribute to *y*'s well-being is that *y* eats an apple. Suppose *x* cares about *y* and suppose *z* has an intrinsic desire that *y*'s good be realised. Surely, *x* and *z* would both want that *y* eats the apple. Darwall, however, thinks that, in addition to this direct object, caring involves an “indirect object”. But what *x* would be wanting for *y* in addition to what *x* and *z* already both want I fail to see. This point applies to the stranger in the train case. If *z* forms a genuine desire (rather than a simple whim or fancy) that the stranger's good obtains, then I cannot see what *x* (i.e., someone who actually cares for *y*) would be wanting for *y* in addition to what *z* already wants for *y*.

The most problematic point for Darwall's account, however, is his definition of caring for *y* as wanting something for *y* for *y*'s sake. The expression “for a person's sake” seems to be a synonym of well-being or interest. If I act for a person's sake, I do something with an eye to his or her interest. An expression such as: “Do it for my sake!” would also confirm it. It conveys that if the action were not done, the person imploring that the action be done would thereby be made worse off. If this is so, Darwall cannot, at the risk of rendering (D\*) uninformative, appeal to the notion of care on the right hand side of the biconditional.<sup>13</sup> That is, if (D\*) is meant to give a characterisation of well-being in terms of something other than itself then the “for one's sake” clause cannot appear on the right hand side of the biconditional.

### 2.5. Of depressives, self-loathers, and conceptual incoherence.

Darwall argues in favour of the claim that (W) cannot be conceptually true by claiming that when a depressive or a self-loather takes himself not to have reason to desire

---

<sup>12</sup> Darwall, 2002b, 47.

something that he takes to be good for him, he does not show *conceptual* incoherence. At most this person would be substantively confused or, as Darwall wants to say, he would wrongly believe himself to be unworthy of care. Our reply to Darwall takes its starting point from another thesis he endorses, namely, the “internalist requirement”:

unless a consideration is one by which an agent would be motivated were she to take account of it rationally, it can give her no reason to act. To be a reason for her, a consideration must be something *on* which she could (autonomously) act.<sup>14</sup>

I take Darwall’s internalism to express a requirement for normative reasons of this form: “R is a normative reason for S to  $\phi$  only if, had S taken account of R rationally, R would motivate S to  $\phi$ .” If, after taking account of a consideration, I believe that this consideration is a reason for me to act, then I am motivated to act. What we are interested in is not so much the motivational side of this requirement but the “taking rational account” side of it. Take the case of the depressed agent: he takes rational account of the consideration that something is good for him and doesn’t think it is a reason for him to desire and do anything. Starting from this view, I will make four remarks. Together, these remarks should show that the best explanation why the depressive and the self-loather do not see that they have reason to desire their good, does not yield the conceptual possibility that well-being entails no reasons. This explanation would point rather to the conclusion that these agents are rationally defective either in their thinking about their good or, alternatively, that they are subject to a form of practical irrationality.

1. In certain cases depression is a serious mental illness; failure to treat manic depression appropriately, for example, may ensue in suicidal behaviour on behalf of the depressive. It is not clear how serious a mental condition affects Darwall’s ‘depressive’. We know that to him, “the thesis that one’s own good or welfare entails reasons for acting will mock the

---

<sup>13</sup> Given that it is rational care that is supposed to be the independent variable, i.e., the variable explicating well-being. See Darwall, 2002b, 9.

<sup>14</sup> Darwall, 2002b, 40. A quick comment on Darwall’s formulation of internalism is in order. It is made up by two sentences. Someone who disagrees with the first sentence does not have to disagree with the second and someone who agrees with the second sentence does not have to endorse the first. Those who defend the claim that an agent can see reason to act in a way and at the same time fail to be motivated to act accordingly can still accept that “To be a reason for her, a consideration must be something *on* which she could (autonomously) act.” They could consistently do so if they thought that having the capacity to autonomously choosing an act does not always amount to being motivated to choose that act. On this view autonomy would be a dispositional capacity.

truth.” But if the depressive’s condition was serious enough, how much authority should we give to the claims he takes to be mocking the truth? Should we, as Darwall does, take them as important indicators of what conceptual truths there are? The internalism endorsed by Darwall requires that the agent “takes rational account” of considerations of his good. But it is precisely part of the condition of being depressed that one is not able to think rationally about one’s own good. The depressive’s rationality is defective precisely in one crucial respect, i.e., that which would let him see that he has reason to desire what is good for him. That is why we should not consider his conclusions regarding the normativity of his well-being as authoritative.

2. Suppose, for the sake of the argument, that we can trust the result of a depressive’s “taking rational account” of his good. He concludes that his good gives him no reason to desire anything. According to Darwall the depressive is not conceptually confused but (consciously or unconsciously?) he wrongly thinks that he is unworthy of care. The following example should show the falsity of this point and the truth of the claim that an agent’s good is normative with respect to his desires. Suppose that Jack is depressed and does not take considerations of his well-being to give him any reasons. You are a good friend of his and want to persuade him to reconsider his judgement. What kind of facts are you going to point out to him in order to convince him? Rather than showing him that he is worthy of care (whatever that would mean), I think you would begin to list all the various ways in which he stands to be benefited if he stopped staying in bed all day, being depressed. If he came out with you, you would say, you could visit an art museum where a new exhibition of his favourite painter has just opened. Alternatively you may both go for a picnic at the seaside or maybe play football; your team-mates, you go on to say, miss Jack badly both, on the field and outside it. What you are doing is to show that Jack *does* have reason to desire the various things that are good for him simply in virtue of these things’ being good for him. You are trying to make him *see* that he has reason to desire these things, by making vivid to his mind the way in which they are good for him (by being enjoyable, for example). Considerations of worthiness of care would be irrelevant.

3. There are two ways in which Jack could respond to your solicitations: (a) he can see he has reason to desire taking part in these activities and be motivated to do so; (b) he can see that these activities are good for him, take himself to have reason to desire them, but fails to desire them and to be motivated to act on them. Maybe his depression is too serious; he



might need therapy rather than friendly solicitations. What is important here, however, is that the depressive is not conceptually confused, hence, we need not postulate the possibility that he sees something to be good without taking it to be a reason for him. Rather, the agent is simply irrational about his desires and actions.<sup>15</sup> That would not be a surprising result given his psychological condition.<sup>16</sup>

Darwall wants us to envisage the self-loather as someone who takes the fact that a certain state could be described as one in which he stands to be benefited to be a reason not to desire it and not to pursue it. I think, however, that there is a more accurate description of this situation, one that is compatible with the view that (W) is a conceptual truth. When we loathe ourselves, if we are at all to be considered as rational, we take ourselves to do so for some reason or other; we do not simply start conceiving of the things that are good for us as undesirable. We might for example hate ourselves for something we have done and decide to punish ourselves, maybe, so as to redress the situation. At this stage this view offers two alternatives: the fact that I stand to be benefited by something counts as a reason to desire and pursue that thing though this reason is overridden by the stronger reason to punish oneself. Alternatively, the reason giving force of considerations of one's well-being is silenced by the reason to punish oneself. These descriptions of the self-loather, however, make him a perfectly rational figure who is neither conceptually nor practically confused, at least insofar as he is justified in thinking that the overriding or silencing reason he takes himself to have really is overriding or silencing. The self-loather, however, may also be a rationally defective figure. He would be so when, in the absence of opposite (overriding or silencing) reasons, he took the fact that something is good for him to be a reason against the pursuit of that thing. This agent would fail to be affected by a reason giving fact in one of the relevant practical ways.

4. In conjunction with (b), another important type of case, call it (c), can explain away Darwall's conceptual possibility. According to (c), the (depressed) agent in fact sees as

---

<sup>15</sup> The idea of irrationality appealed to here is rather narrow: An agent is irrational if, and only if, he takes himself to have (sufficient) reason to  $\phi$  (where  $\phi$ -ing is any type of act) and yet fails to be motivated to  $\phi$ . As Scanlon puts it: "Irrationality in this sense occurs when a person recognizes something as a reason but fails to be affected by it in one of the relevant ways." Scanlon, 1998, 25.

<sup>16</sup> It should be pointed out that Darwall's internalist requirement would not allow (b). If the agent could see he had reason to desire and do something, he would be motivated to do it. But isn't the case described by (b) an ordinary case of an agent's acting against his better judgement? Jack can see that he has reason to do a certain action but fails to be motivated to do it. Can Darwall's internalist requirement allow for these cases?

good for him the things that those who can think rationally know not to be good for him. In our example, Jack may be substantively wrong about what things are good for him. If the depression is serious enough Jack may be left utterly cold by his friend's listing of all the good things he could take part in. In fact, Jack cannot even conceive leaving his bed as anything other than a very unpleasant option. Maybe he can still remember how much he enjoyed these activities back when things were different. Now, however, it would not be the same; he thinks he knows he couldn't enjoy his time with his friends. He just wants to be left alone: that is what he thinks is best for him now.

We can sum up our argument as follows. Firstly, we should be wary of any project that takes as evidence for the normativity of the concept of well-being the thinking of agents who might very well be impaired in their capacity to think rationally about their good. Our reply to Darwall takes its start from the way people who are not impaired in this way would conceive of well-being. Also, when someone not impaired in this way tries to motivate an agent to desire his good, he will not appeal to considerations of worthiness but to the way in which various things and activities could be good for the agent. Finally, depressives and self-loathers are best understood as practically irrational agents. Thus, the best explanation of the case of the depressed and the self-loather does not yield the conceptual possibility that well-being entails no reasons. This is as much evidence as one could ask to claim that (W) is after all a conceptual truth.

If our considerations are correct, in sections 2.3-2.5 we have reached the following conclusions. Darwall's claim that the normativity of well-being must be seen in relation to worthiness of care as in (D') can hardly be considered as a conceptual truth. Also, the notion of care presupposes that of well-being: for this reason the former cannot be used to define the latter nor can it be used to explain its normativity. Darwall used the case of the depressive and the self-loather to show that it is conceptually possible to think that something is good for an agent and yet think that the agent has no reason to desire it. The truth of this claim would show that (W) is not conceptually true. Our reply to Darwall is that the best explanation of these cases should not postulate this conceptual possibility.

### 3. The reason-to-desire theory and the scope problem

#### 3.1. Rational self-sacrifice and well-being

To this point, we have gone along with Darwall in assuming that rational self-sacrifice and well-being are essentially connected. In the remaining part of this chapter we shall keep doing so though it must be noted here that philosophers such as Joseph Raz and Scanlon do not seem to share this assumption. We shall have an opportunity to discuss their views in Chapter 7. How can the reason-to-desire account deal with the scope problem? Darwall has already pointed out what I take to be the most serious challenge facing our account. According to him, it is possible for an agent to form an intrinsic desire for something the obtaining of which would not contribute to his or her well-being. At this stage, Darwall mentions the sort of associative process Mill appeals to in order to explain the genesis of intrinsic desires for things such as wealth. The scope of desire would simply be too wide to individuate all and only the things that contribute to a person's well-being.

I would agree with Darwall insofar as we do at times form intrinsic desires for things that are not good for us. The question is, however, whether we have reason to desire these things. How are we to show that what there is reason for an agent to desire is by definition something the obtaining of which will contribute to his well-being? I have already offered an argument to this effect in our discussion of desires in Chapters 3 and 4. In Millian fashion, we would say that desiring something involves the tendency to see that thing as good for one to have. Of course, not all things we desire are good for us, and that is why our well-being claims can be considered authoritative only after the epistemic process we sketched in Chapter 4. Next we need to show how the reason-to-desire account side-steps the scope problem. Once again we would appeal to Mill and his distinction between desiring and willing. Whenever we have reason to act on the basis of what we have reason to desire, our actions can be expected to contribute to our well-being. Not all our rational actions, however, proceed from our rational desires in this way. At times we may have reason to will an action without having reason to *desire* to do it. This is precisely what draws the boundary between rational pursuits that are *not* in the agent's interest or good and rational pursuits that most contribute to the agent's well-being.

Darwall may insist: surely there are things one would desire and continue to desire even after reflection, the obtaining of which would not contribute to one's well-being. Here is one example offered by Darwall:

... although concern for a person gives rise to a desire for her welfare, it is not similarly related to a desire for outcomes that *she* desires, even rationally. Suppose your friend Sheila is in the following situation. By donating all her disposable wealth she can realize an outcome she cares very much about, say, the rebuilding of a city ravaged by war to a certain degree,  $D$ . But there is a catch. Sheila also has a degenerative disease, which, if it is not checked, will create memory loss and confusion severe enough so that she will be unsure where her money has gone, unable even to hold stable beliefs about the state of rebuilding in the war-ravaged city. Happily, there is a relatively inexpensive drug that can arrest the symptoms of Sheila's disease without side effects. However, the drug is not free, Sheila will not accept donations, and she cares so much about rebuilding the city that even though the difference the cost of the drug would make in the rebuilding effort is quite small (call it  $d$ ), she nonetheless wants (and would continue to want on reflection) to forgo the drug and donate all she has.<sup>17</sup>

Sheila can choose between two outcomes. One in which the city is rebuilt to degree  $D$  together with her advanced disease, memory loss, and uncertainty about the city actual state. And another involving the reconstruction of the city to degree  $D$  minus  $d$ , her knowledge of her role in the reconstruction, and a generally improved mental state. Darwall tells us that Sheila ranks the first outcome higher than the second one and insists that her ranking would survive reflection by Sheila. Importantly, Darwall thinks that Sheila's interest would be best served by the second option. Hence, we may have a rational desire (a desire that survives reflection) for something that is clearly against our well-being. I would like to make two objections.

Firstly, I would like to question whether Sheila's ranking really is something that would survive reflection. The point is that Sheila's desire to put every last penny to the rebuilding of the city, even when that could have little effect on the city and a great effect on her health, is unreasonable. Another feature makes the example as a whole even less credible. If the reconstruction of the city were what most matters to Sheila and by accepting

---

<sup>17</sup> Darwall, 2002b, 43.

donations she might achieve an even higher degree of reconstruction, would it not be rational for her to accept donations?

Our second objection is more important and takes its root from the claim that Sheila's interest would be best served by the second option. Suppose that Sheila has in fact most reason to *desire* the first outcome. Why assume that that outcome is clearly not the best outcome in terms of her well-being? Darwall must take it that this example captures a consensus of intuitions, that we would all agree that Sheila's well-being is clearly not served by the first outcome. I would argue that that is not as intuitive as Darwall thinks. To show this we need to introduce another piece of argumentation.

### 3.2. The expansion of the self

When we form an intrinsic desire for another person's well-being we would conceive of that person's well-being as part of ours. To say that another person's well-being is part of mine means that how well my life goes for me depends on how well this person's life goes for him. This thought dates back to at least Aristotle who claimed that we can justify concern for one's friend and family members as cases of, or on the model of, self-love. The excellent person, Aristotle thinks, is related to his friend in the same way as he is related to himself, since a friend is another self; therefore, just as his own well-being is choiceworthy for him, his friend's well-being is choiceworthy for him.<sup>18</sup>

Now, thinking back to Sheila's case we can say the following: just as we may identify our good with other people's good, we may identify our good with rational ends such as the reconstruction of the war-ravaged city. John Skorupski notes a problem with this kind of picture:

...when the good of others – or outcomes which (in one sense) do not involve me, such as the long-term regeneration of the Caledonian forest – become 'a part of my good', is this not a dilution of the notion of personal good, a fading of its boundaries, in that the notion of self fades from the content of the desire?<sup>19</sup>

---

<sup>18</sup> See Aristotle, 1984, 1170b 6-9.

<sup>19</sup> Skorupski, 2000, 319.

However, he offers the following suggestion:

Can it be plausible to see it instead as an *expansion* of the notion of the self, an enlargement of its objectification in the world through self-identification with others, or with outcomes beyond my stream of experience? It seems to me that... [one] could plausibly argue that there is both dilution and expansion: the psychological process involved is the fading of a lower, because narrower, conception of the limits of one's self (or more accurately one's good) in favour of a more expansive one.<sup>20</sup>

Skorupski's suggestion seems to apply well to Sheila's case. It is conceivable that Sheila identifies her good with the reconstruction of the city. A little reflection shows that many of us are familiar with the phenomenon of the expansion of the self, though, maybe, not in these philosophical terms. We all, to some extent, identify our lot with that of our beloved ones, our friends, our football team, our country, and our projects. What happens to them has direct repercussions on how well off we are. If we commonly identify our well-being with such entities why deny that Sheila could identify her well-being with the reconstruction project? I believe that rather than excluding it as extraneous, any theory of well-being must either account for this phenomenon or, at worst, explain it away as an error.

Darwall may accept that, considered on its own, success in her rational project would contribute to Sheila's well-being. But the outcome that Sheila prefers includes a diminution in her health so great that the loss in well-being would be greater than the contribution afforded by her success in the project. Surely, however, we reasonably identify with our well-being states in which our health is seriously diminished. Think about the case of a mother who agrees to donate her kidney to save her daughter's life. The mother's health will be impoverished by the loss of her kidney. But the contribution to her daughter's well-being, with which her well-being is identified, is so great that her own well-being would be greater were she to donate her kidney rather than let her daughter die.<sup>21</sup>

---

<sup>20</sup> Skorupski, 2000, 319.

<sup>21</sup> Though I believe the mother to have sacrificed something for her daughter I do not take her action to be an instance of *self-sacrifice*. I will explain the difference between sacrifice and self-sacrifice in Chapter 7, §4.3.

The structure of Sheila's example is not different from this one. If, however, we are still inclined to say that the loss of well-being is greater in the first outcome, then I believe we should say there is no reason for Sheila to *desire* that outcome. On our account, however, saying that does not yet amount to saying that she has no reason to will that outcome. Maybe, say, the belief that the reconstruction of the city would substantially enhance the well-being of thousands of people might motivate her to will the first outcome. Her willing might arise without any intrinsic desire to benefit these people (and with an intrinsic desire to preserve her health). Re-described in this manner, the example clearly captures a case of rational self-sacrifice.

#### 4. Well-being or personal good?

Some may retort that the notion of well-being I have been sketching is simply too broad. Sumner reminds us the concept of well-being (or welfare) must fit a certain description. There is, according to him, a pre-analytic agreement about what areas the concept of well-being can cover. Sumner also reminds us that though this agreement may be unquestioned for a set of core meanings of the concept, it may grow thin at the margins.<sup>22</sup> Darwall makes a remark similar in spirit when he writes:

Of course, we might, as a semantic matter, simply use 'welfare' to refer to whatever is useful in promoting the agent's current desires, her rational desires, or some subset of these. But if we do, we will lose the conceptual connection between welfare and benefit...<sup>23</sup>

As we saw above, Darwall was concerned with separating the concept of 'rational choice' from that of 'rational self-interest'. Railton shared the same concern and attempted to solve it by drawing a distinction between the notion of an 'individual's good' and that of an 'individual's welfare':

---

<sup>22</sup> See Sumner, 1996, Chapter 1.

<sup>23</sup> Darwall, 2002b, 35.

This notion [that of an individual's good] is not the same as that of an individual's welfare, for it may turn out that an ideally informed and rational individual would want to seek as an end in itself (were he to step into the place of his present self) the well-being of others as well as himself.<sup>24</sup>

Unfortunately, though, that an individual would have a rational desire for another's well-being does not in and of itself show that the satisfaction of his want would not be part of his well-being. The other person's well-being might be considered to be a part of his well-being, and in that case it would be desired and aimed at as an end in itself. Yet, there is a more important problem for Railton's strategy. In a similar vein to Darwall's, I would want to say that his move threatens to sever any conceptual link between a person's good and a person's welfare or well-being. It seems hard, at least at a pre-analytic level, to accept that the realisation of the same end may have at the same time opposite effects on a person's good and on a person's well-being. And yet accepting Railton's proposal would entail exactly this consequence. Suppose that  $x$ 's choice to promote  $y$ 's well-being involved a considerable loss in  $x$ 's welfare, and yet that choice was determined by  $x$ 's ideally informed desire. On Railton's account the realisation of  $x$ 's end would contribute to  $x$ 's good but would diminish the measure of  $x$ 's well-being. That seems counterintuitive. It is hard to conceive how an individual's good may be positively affected while, at the same time, his well-being is negatively affected. A semantic decision such as this one should not be allowed as it clashes too strongly with our ordinary intuitions.

What are the limits of the concept of well-being, then? And does the notion of well-being we are defending here outstretch them? It might be very hard to give a definitive reply to this question. Philosophers' intuitions about what things are covered and what things are not covered by the concept of well-being (and concepts akin to it, i.e., happiness, interest, good, etc.) have been clearly divided, at least at the margins, since Aristotle and even before. However, to find a pragmatic solution to our problem it might be helpful to look at Shelly Kagan's discussion of this question.<sup>25</sup> Towards the end of his essay, Kagan makes the following point:

---

<sup>24</sup> Railton, 1986, 30 n. 9.

<sup>25</sup> See Kagan, 1992.



Increasing well-being is providing an intrinsic, ultimate benefit to the person; thus it would have to involve altering the person's intrinsic properties. Since a person just is a body and a mind, changes in well-being would have to involve changes in the person's body or mind. ... If something is to be of genuine (ultimate) benefit *to* a person, then it must *affect* the person; it must make a difference *in* the person. That is it must affect the person's intrinsic properties. Changes in merely relational properties cannot be what is of ultimate value *for* the person. I certainly have no argument for this claim. It is simply that I find it overwhelmingly plausible.<sup>26</sup>

Yet, only a little later, Kagan notices that many people will find it overwhelmingly plausible to say that a person who has been deceived into thinking that she is loved and successful, is not as well off as she would have been had she been genuinely loved and successful. On Kagan's model, this would then be an example of a change extrinsic to the person capable of affecting the person's well-being. On our theory a person's well-being would include what Kagan calls 'external goods', i.e., goods that do not affect a person's (body and mind) intrinsic properties. An agent may very well have reason to desire to have genuine friends. The fact that, unbeknownst to him, what he believes to be his friends are not genuine can have a negative effect on his well-being. On Kagan's account, goods such as this one would not be part of a person's well-being. Yet, Kagan adds:

... even if this is so, this does not at all show that the various external goods are not genuine goods. Indeed, they may well be more significant than well-being itself. We will still care deeply about the presence of these external, relational goods. And nothing suggests that we are mistaken to do so. Even if they are external to well-being, there will remain an important sense in which these goods are *personal*, in that their value lies in their relation to the given person.<sup>27</sup>

If one were to grant Kagan his appeal, then the term *personal good* would be most appropriate in capturing all of the things that there is reason for an agent to desire. In that case, then, the goods making up one's personal good would be both those 'internal' goods contributing to the restricted notion of well-being plus other external goods. Notice, however, that this semantic decision would fall prey to the same objection we raised against Railton's distinction between 'individual's good' and 'individual welfare'. Suppose

---

<sup>26</sup> Kagan, 1992, 186.

that suddenly and behind my back, my good friends became false friends and, at the same time, resolved to deceive me about their genuineness by pretending to be even nicer friends than they used to be when they were genuine friends. My personal good would be decreased by my loss of genuine friends though my well-being would be likely to increase through the extra niceness of my (newly false) friends. If the loss in external good is large enough, my personal good would decrease overall though my well-being would increase. Just as for Railton, this solution seems unnatural.

I do not think that we should worry too much about Kagan's distinction between internal and external goods and the way both kinds of goods relate to 'well-being'. That is because Kagan has not convincingly shown that the meaning of well-being has to be restricted to what he calls internal goods. His case was argued for through an appeal to an intuition. There is nothing wrong with appealing to intuitions, were it not the case as it is here, that opposing intuitions are just as strong. In what follows, then, I will take it to be semantically permissible to use 'well-being' in the way we have so far used it. In fact, I think we would even be allowed to use the expression 'personal good' as a synonym for it.

## 5. Posthumous well-being

A discussion about the scope of 'well-being' would be incomplete without at least some remarks concerning its most paradigmatic case: can posthumous states affect one's well-being? Post-mortem states motivate us to act, whilst alive. Their content might be clearly connected to what an agent takes to be his good. Posthumous fame is a typical example. Many people desire to be remembered in history. That one's offspring prosper is another much more common example. Are these irrational desires to have? Or is it wrong to conceive of them as desires under the idea of one's good?<sup>28</sup> I think on our account an agent might very well have reason to desire things that will only obtain after her death. We would have to defend, then, the view that post-mortem states can contribute to a person's well-being.

---

<sup>27</sup> Kagan, 1992, 188-189.

<sup>28</sup> Sidgwick, for example, noticed that people desire things such as posthumous fame though he thought this desire to be irrational. Sidgwick 1907, 113-114, 128-129, 398-402.

Suppose that ensuring a prosperous future for his children is an important part of Henry's well-being and one of his main aims in life. Had he not had such an end, he would have taken up a less stressful but less remunerative job, retired earlier, spent much more money on himself and much less on his children's education. Yet, being a rather late parent, he is perfectly aware that his chances to be around when his children will be able to harvest the fruit of his labour are very small. And so it happens. Good old Henry passes away and only a good few years later do his children fully enjoy and appreciate what he had done for them. However, for the sake of the ensuing discussion, consider an alternative scenario: Henry passes away and a few days later, his children are killed in a tragic car accident.

I hope many will join me in thinking that Henry's well-being is affected in both scenarios though, of course, in contrasting ways. If one does accept this as a possibility then one must think that the following view is true:

(PM) Something can determine the level of an individual's well-being even after the individual's biological death.

(PM), however, seems to clash with another view, which, I believe, also enjoys our intuitive approval. Let us call this view the "Existence Requirement":

(ER) Something can affect an individual's well-being at  $t$  only if the individual concerned exists at  $t$ .

One easy way to dismiss (PM) is to argue that in order for something to affect an individual's well-being, the individual must experience/be aware of that thing and, we will assume, if one is not alive one does not have the capacity to experience or be aware of anything. But this criterion seems too strict as it does not allow cases such as the deceiving friends we mentioned in the last section, to have an effect on an individual's well-being. As that example showed, changes in our circumstances may affect our well-being even though we are not aware of them. But there is another step to be taken to allow that changes we cannot be aware of because we are dead can affect our well-being. In the example of the deceiving friends there is someone whose well-being is being affected. In

\$the post-mortem case, the subject whose well-being is allegedly being affected no longer exists or, more precisely, is no longer alive. That is precisely the intuition behind (ER).

One solution to preserve the spirit of both (PM) and (ER) is to say that posthumous events can impact retrospectively on a subject's life, affecting its success or failure. George Pitcher writes:

[T]he sense in which an ante-mortem person is harmed by an unfortunate event after his death is this: the occurrence of the event makes it true that during the time before the person's death, he was harmed—harmed in that the unfortunate event was going to happen. If the event should not occur, the ante-mortem person would not have been so harmed. So the occurrence of the post-mortem event is responsible for the ante-mortem harm.<sup>29</sup>

Applying this thought to Henry's example, we get that whatever happens to Henry's children after his death, i.e., whether they die in the car accident or whether they live to prosper, will affect Henry's well-being before his death. This solution, however, is incompatible with our account. That is because, on our account, a person's well-being is nothing over and above the attaining of (a particular arrangement of) the various constituents of his well-being. Now, Henry has reason to desire while alive at  $t_1$  that a certain state obtains at time  $t_2$  after his death. If that state is realised at  $t_2$  Henry's well-being will be contributed to at  $t_2$  (if at all) and not at  $t_1$ . At  $t_1$  there is a lack of the most essential thing: the good the attaining of which would contribute to Henry's well-being.

The solution I am rather more inclined to endorse to preserve the spirit of both (PM) and (ER), has been offered by David Ruben. In a footnote he writes:

If one regards  $x$ 's being harmed by  $y$  as a real change in  $x$ , the relatum which is harmed, then it will follow that one cannot harm the dead. On the other hand, if  $x$ 's being harmed by  $y$  is a Cambridge change in  $x$ , then there is no reason in principle why one cannot sometimes do things which do harm the dead.<sup>30</sup>

<sup>29</sup> Pitcher, 1984, 187-188

<sup>30</sup> Ruben, 1989, 232-233 n. 20. The following examples illustrate the difference between real changes and Cambridge changes: "A real change is a change as ordinarily understood: The change in a schoolboy if he comes to admire someone he did not admire before, the change in a woman when she gives birth to a sixth child, the change in an object when its color changes. Cambridge changes are phoney qua changes: The change in Socrates when a schoolboy comes to admire him, the change in the number six each time it ceases

If we bring the results of this view to bear on (PM) and (ER) we get the following. Something can be a *real* change in an individual's well-being at  $t$  only if the individual concerned exists at  $t$ . However, something can be a *Cambridge* change in an individual's well-being even after the individual's biological death. Note, however, that on this account it is also possible that whilst alive, our well-being can be Cambridge affected. Think back to the "deceiving friends" example. Suddenly and unbeknownst to me, my friends cease to be genuine and become deceivers. My well-being is thereby diminished.

If we accept that Cambridge changes may affect a living individual's well-being then we should ask in what way the fact that the existence requirement is fulfilled in this kind of example is relevant in determining that the agent's well-being is being affected. It seems that it is only relevant as a pre-condition: without a person there is no one's well-being to be affected. Note however, that we can no longer say "without a *living* person there is no one's well-being to be affected." To say that would entail that Cambridge changes (of any kind, not only the one's affecting one's well-being) require that what is being changed is alive at the moment at which the change occurs, and that is surely false. We can then reformulate (ER) and endorse it as follows:

(ER') Something can be a *real* change in an individual's well-being at  $t$  only if the individual concerned is alive at  $t$ . Something can be a *Cambridge* change in an individual's well-being at  $t$  only if the agent is alive at  $t$  or was alive at a time before  $t$ .

## 6. Conclusion

Taken together, many of the issues discussed in this chapter lead to one crucial conclusion: a person's well-being can be contributed to by things as diverse as the well-being of other persons, all sorts of projects the outcomes of which (in one sense) do not involve the

---

to be the number of someone's children, the change in Adam and Eve each time they acquire a new descendant." Ruben, 1989, 223. Ruben emphasises the point that though Cambridge changes are always relational, not all relational changes are Cambridge changes.

person, and even states of affairs that obtain after the death of the person whose well-being they contribute to. These are all things that agents may have reason to desire. This conclusion was reached by taking on board an expansionist view of one's good, a view that (a) is in accordance with the phenomenology of our ethical life, and that (b) is able to avoid falling prey to the scope problem. The upshot is that the concept of well-being we have so far described can be very wide in scope, so wide that in fact it may legitimately be asked whether we can still call this concept 'well-being'. I have argued that we can or, alternatively, we can use 'personal good' as a synonym. Given that the limits of the concept are so wide it would come as no surprise to see, as we shall soon do, that many different types of substantive goods can contribute to a person's well-being.

## Chapter 6      The Parts of Well-Being: Some Remarks

### 1. Introduction

Given certain facts about the nature of human beings in general, I will put forward a list of goods that there is reason for a human being to desire under the idea of his or her well-being (§2). It is not an aim of this chapter to argue extensively in favour of each of the candidate goods appearing in the list. Rather, I will assume the plausibility of each of the elements and discuss at greater length what I consider to be particularly tricky issues concerning some of them. It will become clear from our discussion that for any human agent, the range of goods captured under the idea of his or her well-being might be a rather extensive one. After a few brief remarks concerning the list of goods in general (§2) and some of the goods in particular (deep personal relations, virtue, knowledge, and achievement) (§3), I will dedicate one section each to specific issues concerning respectively pleasure and its relation to happiness (§4) and autonomy (§5).

### 2. The list

I would think that the list of goods that there is reason for human beings to desire would include the following ends: pleasure, achievement, autonomy, deep personal relations, knowledge and understanding, and virtue.<sup>1</sup> The question will of course arise: how can we devise such a list? First, it would be a good idea to demarcate the realm of knowledge within which the devising of the list would take place. The investigation would be one in philosophical anthropology. The investigator would have to look at the relevant empirical facts concerning human beings in general and thence come up with some generalisations. What would be the best evidence for the claim that something is desirable for a human being? Take what human beings generally desire as a primitive criterion and move from

---

<sup>1</sup> For a more in-depth discussion of some or even all the items on the list see Finnis, 1980, esp. Chapters 3-4; Griffin, 1986, Chapters 1-5; A. Moore, 1991; G.E. Moore, 1903; Skorupski, 1989, Chapter 9.

there using the epistemological process for normative claims we discussed in Chapter 4. Of course, in everyday life situations human beings will not generally desire pleasure, autonomy or achievement, etc., so much as the various particular objects that bring them about. It is the work of the philosopher to be able to abstract from a number of reasonable desires for particular objects or states of affairs, and conclude that, for example, they all share the feature of being desired under the idea of respectively pleasure, knowledge, autonomy, etc.<sup>2</sup>

It is also important to understand the nature of the list. The list is objective in the sense that given any human being it would be intelligible for him to desire any or all of the items in the list as part of his or her well-being. This claim, however, has to be accompanied by one remark. As discussed in Chapter 4, though it would be intelligible to desire all of the items on the list, it might not be best for any particular agent to do so on one occasion or, if his circumstances are peculiar enough, to ever do so. That will depend on the circumstances of the agents and on certain facts about him or her.

### **3. Deep personal relations, virtue, knowledge, and achievement**

#### 3.1. Deep personal relations

Under the idea of deep personal relations, I intend to group relationships as diverse as friendship, parental love, as well as romantic or passionate love. An in-depth examination of each kind of deep personal relation and the precise way in which they can be pursued under the idea of one's well-being would take a book-length essay and cannot be taken up here. As explained above, such an investigation would belong to the realm of philosophical anthropology. Most abstractly, however, we shall say that all of these types of relations

---

<sup>2</sup> Important substantive questions would be raised if the best philosophical-anthropological investigation would for example exclude some of our entries and include others. It is not the task of this thesis to carry out such an investigation. The list I put forward can be considered as a plausible list on the basis of some of the relevant literature. I will leave it an open question, and a very interesting one at that, whether the list I have offered is indeed the best account given our present knowledge. One particularly interesting issue, for example, is whether dignity should also appear in the list as an independent entry.



involve a special attachment of a being to another.<sup>3</sup> Most importantly, this attachment involves the kind of identification of a being's well-being with that of another, the kind of extension of the self we discussed in Chapter 5.

### 3.2. Virtue

Recall the difference we introduced in Chapter 3 between the *virtuous agent* on the one hand and the *conscientious agent* on the other. The former would *desire* to act virtuously; the latter would do so but out of a sense of duty, and not out of an intrinsic desire. Let us also remember that these are ideal-type figures; the very same agent may combine the two types. One may, for example, act out of a sense of duty on some occasions but find that he has reason to desire to act virtuously on other occasions. Now, I would claim that the virtuous agent is a lover of virtue in a way that is different from that in which the conscientious agent is a lover of virtue. When we say that the virtuous agent loves acting virtuously, we say that she has an intrinsic desire for it. In this respect, acting virtuously involves her affectively and first-personally. Not being able to act on her desire would translate into a genuine loss of well-being for her. In this sense, her action would not be selfless. That, however, should not be understood as meaning that her action would be egoistic or selfish. The process is once more that of the expansion of the self. Her well-being is to some extent identified with that of the being she is concerned with or with the virtuous action or activity itself. The frustration of her desire to act virtuously would be perceived as a personal loss. This, I think, should be a familiar enough phenomenon of our ethical life though probably not as recurring as the expansion of our well-being towards our loved ones.<sup>4</sup> Things are different for the conscientious lover of virtue. He will act virtuously entirely out of his sense of duty.

---

<sup>3</sup> Though necessary, this condition might not be sufficient to capture, for example, a relation such as friendship. Two people can't be said to be genuine friends unless the special attachment in question is reciprocal (can we be friends if you despise/hate me?). But, then again, friendship is most typically a social form. What particular practices are involved by it will depend to some extent on the particular society in which it is instantiated.

<sup>4</sup> Note the difference between my account of the virtuous agent and Mill's. On my account, acting virtuously is not to be understood under the idea of pleasure as it was for Mill (though it does not exclude that it might on occasions be so understood). Rather, acting virtuously is an independent categorial end under the (non-substantial) *idea* of one's well-being.

### 3.3. Knowledge

In what way can knowledge or understanding of one's situation be taken to be part of a person's well-being? Take the case of the mother whose son has long gone missing in a military campaign. She wants to know his fate. She does not desire to know because she is miserable not knowing; she is miserable not knowing because she desires to know.<sup>5</sup> Knowledge of one's situation is an end that people do desire and, it seems to me, they do so with good reason. Importantly, as this example shows, they can desire it as a component of their own good and not for the good of others or in pursuit of an ethical ideal. An illustration of another feature of how knowledge could be good for a person is given by the following example from Robert Nozick:

We would not wish for our children a life of great satisfactions that all depended upon deceptions they would never detect: although they take pride in artistic accomplishments, their critics and their friends too are just pretending to admire their work yet snicker behind their backs; the apparently faithful mate carries on secret love affairs; their apparent loving children really detest them; and so on.<sup>6</sup>

It would be hard to judge as good for them the life of people not knowing the grim truth about their situation. Many, and I am inclined to join them, would agree that these people have reason to desire to know the truth about their situation, to stand in the correct epistemic relation to the facts. Needless to say, knowledge has great instrumental value. But standing in the correct epistemic relation to the facts can in itself be good for one. We desire to have *genuine* friends and not simply ones that seem so; we desire to accomplish great artistic feats, not merely to have the false experience of having done so. More generally, we want our beliefs to be based on what is actually the case. Take away this feature from our beliefs about our experiences and many of us might have good reason to stop desiring having these experiences altogether. This, however, should not be taken to imply that knowledge and understanding are desirable only as a part of a Moorean organic unity. I do not agree with G. E. Moore when he writes that:

---

<sup>5</sup> I borrow this example from Skorupski, 1989, 300.

<sup>6</sup> Nozick, 1989, 105-106.

Knowledge, though having little or no value by itself, is an absolutely essential constituent in the highest goods, and contributes immensely to their value.<sup>7</sup>

Take the case of the mother once again; the experience of coming to know the fate of her son might be very painful and certainly undesirable in that respect. Yet, she desires to know. Would knowing about her son be significantly valuable for her only when contributing to a greater unity? I don't think so. Take also the nature of scientific discovery. Many human beings display a primitive curiosity about the natural and social world around them and their place in it. They seek this knowledge for its own sake. They do not seek it only to receive the pleasure ensuing from the satisfaction of their curiosity. Nor do they seek it because the discovery process is overall a pleasurable one or because coming to know is pleasurable. Many, I suppose, would want to satisfy their curiosity irrespective of any pleasurable state that might be involved in or ensue from the process.

#### 3.4. Achievement or accomplishment

Some authors want to distinguish between achievement and accomplishment. Andrew Moore, for example, takes it that:

one accomplishes something, in the prudentially important sense, if and only if one makes a substantial contribution to a project or end which is itself of major value.<sup>8</sup>

Many activities and projects lack the value necessary for accomplishment. The cracking of 30 walnuts in 57 seconds between one's buttocks, for example, is some achievement, but no accomplishment.<sup>9</sup>

Moore's terminological choice seems to be stipulative. I do not mean to imply that there is something wrong with it; I only mean that we might have inverted the terms without offending against our pre-analytic intuitions about their usage. The substantive underlying issue, however, is more important. There are *doings* that contribute to an agent's well-

---

<sup>7</sup> Moore, 1903, 199. Note that what my exposition is neutral on the question concerning the plausibility of organic unities, i.e., unities the values of which might be much greater than the value of each of its components taken together. Recently, Zimmermann has argued against its plausibility. See Zimmerman, 2001, Chapter 5.

<sup>8</sup> Moore, 1991, 212-213.

<sup>9</sup> Moore, 1991, 216.

being that are “of major value” and *doings* that are not “of major value”. I think there are two interesting issues to be addressed: (1) whether the second kind of doings can be good for the agent at all, and (2) whether accomplishing and achieving have some core meaning in common. Let us start with the latter point.

Both kinds of *doings* seem to share a central feature, which we may call *agency authorship*. Let me illustrate what I mean by that. Sophie might rightly think it very important that the truth about the origin of the human species be discovered. However, she does not care about being herself engaged in any of the scientific projects conducted to that purpose. In fact, she knows herself well enough to realise that she does not have the right kind of disposition to become a professional researcher in bio-archaeology. Researching the truth about the origin of the human species is not an authorship-centred aim of Sophie. However, one of Sophie’s most important aims is that of climbing the ten highest mountains in the world. She does not care about other people having already done so. Nor does she care about having the mere experience of climbing the ten mountains by being plugged into a powerful simulating machine. She desires to *do* it and she desires that *she* does it. Whatever one’s intuitions about the value or worth of these or any other *doings*, I believe accomplishment and achievement have agency authorship as their common kernel in the way just illustrated.

The point we are left to examine, now, is whether an authorship-centred doing of no “major value” may ever be good for the individual engaged in it. Take Rawls’s famous grass counter example. On our account, any end an individual desires and pursues as a part of his well-being must be subsumable under an intelligible categorial end. Now, we do not find it intelligible to say that counting blades of grass can be desirable for an individual as an end in itself, let alone as an authorship-centred end. If we cannot see something as an intelligible end it is simply hard to see how it could be contributing to a person’s well-being. However, if, say, counting blades of grass were a way to find peace of mind then we would be able to place it under the idea of pleasure and it does make sense to say that pleasure contributes to a person’s well-being.<sup>10</sup> But, then, enjoyment or pleasure would

---

<sup>10</sup> Note that though the activity of counting blades of grass is now placed under the categorial end of pleasure it would still also be an authorship-centred doing. Though counting the grass is constitutive of pleasure, it also matters to the agent that *he* counts grass blades, not that grass blades be counted. Nor would the agent be satisfied with the mere *experience* of counting grass blades. It is part of his aim that he actually *does* the counting.

become the categorial end, and counting blades of grass would only be a *source* and not a part of well-being.

But the question could be pushed further, as, after all, we do not think that all sources of well-being are of equal “value.” The question would then be whether an authorship-centred doing of no “major value” may ever be a source of well-being. Take Rawls’s grass counter again. We have reason to desire enjoyment, but do we have reason to desire enjoyment through the activity of grass counting? Though most of us would find this activity unusual as a source of enjoyment, it is not inconceivable that someone may find grass counting enjoyable in some way. Our intuitions would differ if someone found cutting himself up with razor blades enjoyable. That is because most of us would find that activity particularly painful and what is painful is not pleasurable.

How about Moore’s curious nutcracker-man? Suppose that we are agreed that achievement and accomplishment are a component of agents’ well-being insofar as it is intelligible to want to achieve or accomplish certain things under the idea of one’s well-being. Our question, then, is to understand whether the activity of cracking the most nuts in the least time can be considered as an authorship-centred achievement or accomplishment. If it can then it can be considered as a source of a person’s well-being. We find it easy to say that competitive athletics or finding the cure to AIDS are accomplishments or achievements. We find these activities valuable (here ‘valuable’ is not to be understood prudentially) and we find it intuitive to say that an agent may have reason to desire to achieve or accomplish something valuable. It is harder to find competitive buttock-nut-cracking a valuable achievement or accomplishment, and if it cannot be considered as such an agent cannot have reason to desire it as a source of well-being, not simply because it does not seem as a valuable end but, more radically, because it does not seem to be an intelligible activity.

Someone, however, might press us: what is so different between aiming at being the world’s 100 metres champion and aiming at being the world fastest nutcracker? Can we exclude the possibility of a society in which competitive buttock-nut-cracking is a highly recognised accomplishment or achievement? It might be thought that what kinds of pursuits we can find intelligibly valuable (and ‘valuable’ here is not to be understood only in terms of well-being) depend on our social forms and that we do not find buttock-nut-cracking valuable because it is not a recognised social form. Joseph Raz takes social forms

“to consist of shared beliefs, folklore, high culture, collectively shared metaphors and imagination, and so on.”<sup>11</sup> He then goes on to claim that:

A person’s well-being depends to a large extent on success in socially defined and determined pursuits and activities. ... What are the reasons for [this] thesis? It is not a conventionalist thesis. It does not claim that whatever is practised with social approval is for that reason valuable. It says that the comprehensive goals a person finds valuable are based on social forms, whether or not these are socially approved social forms. In other words the thesis merely sets a limit to what comprehensive [goals] can be valuable for any person. They can be valuable only if they can be his goals and they can be his goals only if they are founded in social forms.<sup>12</sup>

I agree with Raz’s objection to conventionalism. However, we have to understand what exactly Raz means by his last claim. Raz may be taken to say that the goodness of a goal is dependent on its being instantiated in a social form. I disagree with this claim if we understand it to mean that what warrants our judgement that something is good is the fact that it exists in a social form. As we claimed when presenting the buck-passing account of value in Chapter 3, different things are valuable in very different ways. To judge that running the 100 metres in the least time is an achievement or accomplishment is to say that this activity has a feature that provides agents with a reason to *admire* it. We could of course wonder why we find this activity admirable but not buttock-nut-cracking. I would say that it is a primitive fact about human beings that we *spontaneously* find some things admirable and not others.<sup>13</sup> Of course that alone is not sufficient to warrant our judgement that something is admirable. The consensus of competent judges would have to be appealed to as an indicator that our primitive normative impulse is indeed authoritative. Going back to Raz, we should understand the dependence of valuable activities on social forms as follows: the existence of the latter would provide an opportunity to instantiate the

---

<sup>11</sup> Raz, 1986, 311.

<sup>12</sup> Raz, 1986, 309-10. Note that in the second square bracket, I took the liberty of inserting the word “goals”. In the original Raz writes “forms” in its stead. There are two reasons to believe that that was an oversight. Firstly, in the sentence before the one containing the alleged oversight, as well as in the preceding discussion Raz keeps referring to “comprehensive goals” and not “comprehensive forms”. Secondly, in the sentence following the one in question, the subject “they” can only refer to a person’s “comprehensive goals”. If we took it to refer to “comprehensive forms” the second part of the sentence would read as follows: “the comprehensive forms can be his goals only if the comprehensive forms are founded in social forms.” That would not make much sense.

<sup>13</sup> This could be compatible with a genealogical explanation of why running fast and physical strength in general might originally have been found useful and consequently admirable. The same story could hardly be told about buttock-nut-cracking.

former, though the former would be valuable whether or not there existed a social form to instantiate it.

#### 4. Pleasure and happiness.

The main aim of this section is to understand the relation between pleasure and happiness, and their importance as constituents of well-being. The term ‘happiness’ has so far been largely absent from our account of well-being and that may seem odd for a book-length account of well-being such as this. The term ‘happiness’ has different meanings in philosophical discussion and outside it. This fact, I believe, is the source of some confusion, at least in moral theory. In what follows I bring to the fore three important ways in which the term has been used in the philosophical debate. On the one hand, Classical Utilitarianism typically referred to happiness as to pleasure and the avoidance of pain. On the other hand, as in the case illustrated by our discussion of Sumner in Chapter 4, happiness is appealed to as the positive cognitive and affective endorsement of one’s situation and/or life. More recently, however, happiness has entered ethical discussion as a mood or feeling of joyfulness or high-spiritedness.

##### 4.1. The meanings of happiness

Mill famously thought that:

By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.<sup>14</sup>

Mill also thought that happiness, and thus pleasure and the avoidance of pain, is the only thing desirable as an end. It would not be wrong to say that Bentham and Sidgwick, each with their own account of pleasure, would concur with Mill in thinking that happiness is pleasure and the avoidance of pain and is what is ultimately desirable.<sup>15</sup> There are,

---

<sup>14</sup> Mill, 1998, 55, 112.

<sup>15</sup> Though Sidgwick would be more careful in dealing with the question of whose happiness (an individual’s own or that of all individuals) is the ultimately desirable end.

however, two other important senses in which the term 'happiness' has been used in philosophy. James Griffin, for example, writes that:

In a very common use now, to be 'happy' is to be satisfied with or contented with having a good measure of what one regards as important in life. In this use of the word, 'happy' has to do with one's situation; one is fortunate. It also has to do with one's state of mind; one is glad or cheerful. It typically has to do with both situation and state of mind (one has the latter because of the former), but the two elements can appear in very different proportions in different cases. At one extreme, a martyr can go happy to the stake, merely secure in the conviction of right. At another extreme, a person can be happy (cheerful) for a few moments before realizing how unfavourable the situation actually is.<sup>16</sup>

According to Griffin, one kind of happiness ascription refers to a situation in which one is satisfied or contented with what one regards to be important in one's life. This for example is the way Sumner uses the term. One is happy if one believes that one has a good measure of the things that one wants from life. The other kind of ascription refers to certain types of feelings. In this case, one is happy if one feels cheerful, glad, or joyful. Now, the practice of these two kinds of ascription may come apart, as one may be contented with his situation without feeling happy and, *vice versa*, one may feel happy without believing that one has a good measure of the things one wants from life. For the sake of clarity, let us refer to being satisfied with one's lot as *situational happiness* and to feeling happy as *mood happiness*. Typically, the former kind of ascription focuses on longer stretches of time, while the latter sense focuses on shorter ones.

In what follows, I shall not discuss *situational happiness* and the kind of prudential good it could be.<sup>17</sup> We shall rather look at the way in which *mood happiness* is related to the classical utilitarian notion of happiness as pleasure (or enjoyment or pleasurable consciousness). This comparison will be important in order to understand the value of mood happiness as part of our well-being.

---

<sup>16</sup> Griffin, 1988, 226.

<sup>17</sup> The affective state (the feeling of satisfaction or fulfilment) that principally constitutes this type of happiness does not seem to be a good worth pursuing for itself. Rather, in normal circumstances, this feeling would arise as a consequence of an agent's realisation that he has attained goods such as enjoyments or pleasures that are intrinsically worthy.



#### 4.2. Mood happiness and pleasure happiness

We can start by saying that Classical Utilitarians such as Bentham, J. S. Mill, and Sidgwick did not concentrate on the difference between feelings such as cheerfulness, gladness, joyfulness, high-spiritedness, etc. on the one hand, and enjoyment and pleasure on the other. Both moods on the one side and pleasure on the other are essentially phenomenal states. They all involve a feeling, something that it is like to be in that particular mood or something that it is like to enjoy or feel that pleasure. However, moods differ from pleasure or enjoyment insofar as the latter are object-directed and the former are not. I normally enjoy something, but I do not necessarily have to feel cheerful *about* something. I may feel cheerful full stop. The question we would like to ask now is the following: Did Classical Utilitarians ignore something important from the point of view of well-being when they failed to explicitly include these moods in their meaning of happiness? How important in one's life is being in a happy mood, i.e., being cheerful, joyful, high-spirited, etc.?

To answer that question we might want to distinguish between the *causes* of happy moods such as joyfulness, cheerfulness, etc. and the happy moods themselves. Some people are made happy by very simple events. The fact that someone smiles at me at the newsagent when I buy the newspaper in the morning may cheer me up for the rest of the day. Other people are caused to be in a happy mood by much more complex events, such as, say, having achieved the writing of a long paper. More importantly, some of the causes of happy moods are not necessarily to be experienced as pleasant by the subject. The writing of an article may not itself be pleasurable. And yet, it might cause me to be cheerful. Even more, some people may feel happy as a result of no apparent experienced event at all (today I woke up feeling happy). Finally, one may feel happy at the idea of something yet to happen (the idea that tomorrow I go on holiday may make me feel happy now).

Thus, the sources of one's happiness in this mood sense need not themselves be pleasant. Similarly, not every pleasurable experience will make one happy in this mood sense, i.e., cheerfully, joyful, etc. I might find eating an ice cream pleasant and yet fail to be moved in any way. Yet, as a psychological generalisation about human beings, it seems true that pleasant and unpleasant experiences can be singled out as likely sources of happy and unhappy moods. Suppose that I am enjoying walking on the hills. The experience is likely

to make me feel joyous or cheerful. Similarly, suffering intense pain for a protracted period of time is likely to depress me.

I believe that happy feelings, such as cheerfulness, joy, etc. are not things we have much reason to desire for themselves. Rather, these feelings seem to be the product of activities or states that we seem to pursue or avoid in and for themselves. It is because I enjoy hiking that I feel happy when I do so, and not the contrary. Also, as pointed out above, I might very well enjoy that activity without thereby being made joyful by it. More than anything else, the role of mood feelings of happiness seems that of signalling how well an agent is faring at that precise moment in his life. The Classical Utilitarians' neglect to give much weight to these feelings as important elements in one's well-being does not seem to be an important neglect.

The utilitarians were also right in stressing the importance in one's life of enjoyments and pleasures. That is especially so if we take on board Mill's distinction between higher and lower pleasures. It seems that many human agents are capable of enjoying or finding pleasurable different kinds of more or less elaborate activities, achievements of all kinds, and even virtuous actions. What I am not quite clear about, however, is how pleasure and enjoyment would relate to the mood feeling of happiness itself. Is pleasure/enjoyment just the typical cause or source of mood happiness or is itself part of the meaning of happiness? Take an example: suppose I am enjoying eating an ice cream. What I am in doubt about is not the claim that pleasure is a central element of well-being, but rather whether Mill and Sidgwick were right in taking pleasure to be part of the meaning of *happiness*. Am I happy every time I enjoy some pleasure? Or is happiness a term more closely related to feelings such as joyfulness and cheerfulness that might fail to arise even when I enjoy some pleasure? This, in the end, is a terminological issue and from an ethical point of view, not much hinges on it. If happiness is understood to include pleasure then happiness is an important element of a person's well-being. If happiness is understood only in the sense of the mood feelings then happiness is a much less important part of a person's well-being. In any case, pleasure remains an important component of well-being.

#### 4.3. A possible disagreement on moods

Daniel Haybron writes:

... such positive emotional conditions as a predominance of joyfulness, high-spiritedness, peace of mind, etc. would exemplify happiness, while a predominance of their negative counterparts—depression, anxiety, fear, anger, feelings of discontent, etc.—would typify unhappiness.<sup>18</sup>

However, on Haybron's view:

To be happy... is not necessarily to *feel* happy. A generalized low-level positive mood or sense of tranquillity might suffice for being happy without predominantly, or ever, involving the acute emotion of feeling happy. Happiness thus conceived is not a particular emotion or mood at all, but consists rather in a subject's overall emotional or mood state.<sup>19</sup>

If “[t]o be happy is not necessarily to *feel* happy”, then, contrary to what I claimed above, mood happiness or the feeling of happiness is not an essentially phenomenal state. Haybron thinks that happiness consists in a person's overall emotional state: insofar as one's emotional state is basically positive, one is happy. However, there are some perplexing questions about Haybron's passages above: what is a “generalized mood” or “overall emotion”? How would it differ from one that is not so general? What are we to associate with “the acute emotion of feeling happy?” Whatever answers Haybron might want to give, note that, according to him, there are instances in which a person is happy irrespective of his feeling happy and instances where the state of happiness involves an “acute emotion of feeling happy”. It is at this point that my view clearly parts with his. That is because of an apparent discrepancy in our respective conceptions of what the correct epistemology of (mood) happiness ascriptions should look like. It seems to me that if one is happy, then one is in a position to know that one is happy. But how can anyone know that one is happy unless one feels happy? And, what is more, what other supplementary evidence beside the fact that one feels happy would one need in order to judge correctly that one is happy?

Going back to the last quotation, I would want to think about the agent who, according to Haybron, does not feel happy and yet is happy because in ‘a generalized low-level positive

---

<sup>18</sup> Haybron, 2001, 506.

<sup>19</sup> Haybron, 2001, 506-507.

mood or sense of tranquillity'. This agent is in a state not involving the *feeling* normally associated with happiness. The possibility then may arise that the agent is not aware (even after introspection) of his mood. This would be a rather counterintuitive consequence for Haybron's theory, as we tend to think that each person is the best judge of his internal states above all when the state in question is a state such as a feeling.<sup>20</sup> On Haybron's theory, this first person authority would be undermined. Suppose, however, that the agent is aware of his happiness. On the basis of what evidence can he claim that? Suppose that the agent replied that he is happy because he is in a generalised low-level positive mood. The question would then be the same: what makes him say that? I do not see any solution other than that of saying that he feels a certain way, and that that feeling (in this case the feeling of being in a low-level positive mood) is *a* feeling of happiness. Mood happiness, then, is essentially phenomenal.<sup>21</sup>

## 5. Autonomy and its value

I am here concerned with the notion of autonomy and its alleged *value* as a part of someone's well-being. On my definition, autonomy is the capacity to recognise and be motivated by *reasons* and, if autonomy is part of one's well-being as we claim, then there is reason for an agent to desire it. As we shall see, however, what we have reason to desire is not so much autonomy as the capacity itself but the liberty to exercise our autonomous capacity.

### 5.1. The capacity for autonomy

---

<sup>20</sup> Of course, Haybron may simply deny that happiness in his sense is necessarily a feeling. He might say that happiness, at times, is a cognitive state without an affective side and that we might be aware of it cognitively rather than by being affected by it. This would be an over-intellectualised account of happiness and I do not think Haybron is prepared to endorse it.

<sup>21</sup> I might have been unfair to Haybron insofar as his account of happiness is trying to accommodate also happiness in its *situational* sense. In fact, he only *implicitly* notices that happiness might have these two senses, short-term and situational, and seems rather dismissive about the former. That is problematic, however, since he takes himself to be giving an account of happiness in what he calls its *psychological* sense rather than its *evaluative* sense where this distinction is supposed to differentiate the question of what happiness is from that concerning whether happiness should be pursued. However, as argued above, when he deals with something such as *situational* happiness, Haybron is thereby dealing with an *evaluative* sense of happiness.

Traditionally, the understanding of autonomy in terms of a rational capacity (or, as we put it, the capacity to recognise and be motivated by reasons) is associated with Kant. On Kant's view, one is autonomous if one is in a position to subject one's will to self-imposed maxims that conform to the moral law. Our account is clearly Kantian so far as the idea of a "self-rule" goes. The self-imposed rules do not derive from simple urges one feels compelled to act on. To be free is not simply to be able to do what one wants. Our autonomous capacity entails the capacity to resist any motive had one seen good reason to do so. Yet, my account differs considerably from Kant's in three respects. Firstly, I do not claim that every self-governing rule has to conform to the moral law in order for it to belong to the realm of autonomy: there is no necessary link between autonomy and morality. Secondly, autonomy as the capacity is determined by an initial allotment for which one is not responsible. That allotment is a matter of luck and is determined by causes entirely external to the agent. Nonetheless, someone endowed with a sufficient capacity can, through her own effort, extend the exercise of her capacity. Unlike Kant's, my conception of autonomy is not grounded in any transcendental realm. Rational autonomy is a psychological capacity. As a result, and this is the third difference, some autonomous beings are naturally endowed with a greater capacity for autonomous acts and the same individual attains different degrees of autonomy at different stages and moments of her life. Our position, then, does not endorse another distinctively Kantian feature, namely, the idea that all agents must be equal in their possession of autonomy.

## 5.2. Autonomy as a part of well-being: freedom to exercise autonomous capacities

I shall claim that autonomy is a part of our well-being. But the sense in which this claim is true needs to be articulated. Do agents have reason to desire the capacity for autonomy? My guess is that they might but that that is not really a significant prudential value. For one, someone who does not have that capacity in the first place cannot have reason to desire anything let alone a capacity for autonomy. That, however, does not exclude that an agent with a sufficient capacity for autonomy may have good reason to desire to have his capacity *extended* to those areas which could be regulated by his capacity but aren't. But this thought too needs to be made more precise.

We do not extend the *capacity* itself. The capacity has been allotted to us with a given potential. Rather, we may be able to extend the *exercise* of our autonomous capacity.<sup>22</sup> We could then say that, *ceteris paribus*, there is reason to want to *exercise* one's autonomous capacity more rather than less. What's good for one is not the capacity itself, but one's exercise of it. Suppose that I have the capacity to think and act on reasons in several spheres of my active life but that, in fact, I do not exercise my capacity in all of these spheres. The claim would be that *ceteris paribus* I would be better off if I extended the exercise of my capacity to these spheres of activity too. This may be compared to a case in which an agent has the capacity to appreciate certain kinds of pleasure but has never exercised it. Suppose someone has the capacity to find pleasure in the taste of treasured wine but has never exercised this capacity. We may think that there is a *ceteris paribus* good reason for him to want to exercise his capacity for pleasure over treasured wine as well.

Though we can conceptually separate the idea of *exercising* our autonomous capacity from the idea of *possessing* an autonomous capacity, we cannot in practice tell whether a being has the capacity unless it exercised it. Let me demonstrate this point. Agents exercise their autonomous capacity whenever they have an opportunity to rationally consider whether they have reason to act in a certain way. As in Chapter 4, an act is here understood widely, as an epistemic, a practical, or an affective act. Thus, for example, I exercise my autonomous capacity when I consider whether there is reason for me to *feel* a paralysing fear for spiders (given, say, that I live in a place where spiders are innocuous). Or again, I exercise my capacity when I consider whether there is reason for me to abandon my present career based on my desire to try another type of career. Finally, I exercise my capacity when I consider whether there is reason to believe that God exists when I don't seem to have much evidence in favour of this belief. It would be hard to consider as autonomous a being that possessed the capacity for autonomy but that never *exercised* it. This being would never assess his feelings, would never deliberate about his actions, and would never think his beliefs through for himself. In practice, there would be no way for us to tell apart a being such as this one from a non-autonomous one.

---

<sup>22</sup> It is important to note here the underlying debate. According to A. Moore there is a question whether autonomy is a capacity or the exercise of a capacity. He claims that ordinary usage of 'autonomy' goes both ways and that philosophers too go both ways. See Moore, A., 1991, 159, 183 n. 18. In what follows I will never use the term 'autonomy' and always refer to either the 'autonomous capacity' or the 'exercise of the autonomous capacity.'

One reason to think that the exercise of our autonomous capacity is good for us is that it would enable us to better understand what choices and actions are good for us. But this would mean that the exercise of our autonomous capacity is instrumentally good. What we need to understand is whether it is good for us for its own sake. Think about the following example. A few worlds away from this one, a Well-Being Agency is part of the state-funded welfare system. The Agency offers to run the life of those citizens whose well-being persistently falls below the required minimum standard through their own bad choices. Citizens decide whether to have the Agency run their lives for them, for any period of time they wish, free of charge. It is well known that the Agency has an impressive record in bringing citizens' well-being back on track.

Would *we* citizens of Earth find an "Agency-option" attractive? Presumably, the Agency would still leave us the possibility to *exercise* our autonomous capacity. We could still assess our feelings and beliefs, and deliberate about what to do insofar as the Agency thought that that wasn't detrimental to us. Whenever the exercising itself was thought to be bad for one the agency would block it at its inception by means of a sophisticated technological device implanted in our brain. The Agency would then go on to make its own practical choice for us, or implant a certain belief or feeling in us.

Many of us would find this scenario very unattractive. Someone may claim that what we would find unattractive is the feeling of frustration associated with the frustration of our choices. But that is not so. The Agency would block our exercising the capacity at its inception. Our thoughts or choices would not be frustrated, as they would not have the time to arise in the first place. What we find particularly unattractive, then, is the thought that we are not the masters of our own thinking, feelings, and doings. Not to be our own masters in some respect is a loss of freedom. This loss of freedom, however, would be more radical than that of a slave. A slave can still think freely, he could assess his feelings and at least think how he would act were he causally free to act. The loss of control of our own lives would be more radical than that of the slave. To be able to identify with our life as ours, we need to identify with our feelings, beliefs, and actions, and the process by which we identify with them is precisely the exercise of our autonomous capacity. To have an extraneous agent implant feelings, beliefs, and choices in us would be alienating. It would be an intrusion into our body and mind. The loss of the exercise of our autonomous

capacity would amount to a loss of control over our lives but, more radically, to a loss of identity. Mill was right: the individual should be left free to be sovereign in that which concerns his body and mind.<sup>23</sup>

In conclusion, then, to think of the exercise of our autonomous capacity as good for us for its own sake, is to think that we have reason to want to keep our own identity, to have control over the conduct of our own life, our beliefs, and our feelings. The aim of the Agency on planet Earth would be a good one: helping people live lives that are good for them. But the Agency would be likely to fail its mission on planet Earth unless it changed its methods. By (partially) taking away the possibility to exercise one's autonomous capacity the agency would eliminate an important part of our well-being. To understand the value of this part of our well-being is to understand that we have reason to want the liberty to exercise our autonomous capacity. If the Agency intended to increase its chances of success on Earth, it could not force its choices (its thoughts and feelings) on us, but begin to remonstrate with us, reason with us, persuade us, and entreat us whenever it felt our choices to be bad for us.<sup>24</sup>

---

<sup>23</sup> See Mill, 1977, 224.

<sup>24</sup> Once again I am echoing Mill's words. See Mill, 1977, 223-224. Some may think that in this paragraph I have confused valuing the exercise of one's autonomous capacity with valuing the liberty to exercise one's autonomous capacity. There is no confusion. My claim is that having a reason to want for itself the exercise of one's autonomous capacity entails having a reason to want the liberty to exercise one's autonomous capacity.



## 6. Conclusion

One general conclusion can be drawn from this discussion. If I am correct in thinking that such diverse intrinsic goods as the ones listed in this chapter can be parts or constituents of well-being, then contrary to Scanlon the concept of well-being does implicitly or explicitly play a very important role in any human agent's life. If this picture is correct, we may very well have reason to desire many of the things we have reason to do. It will be the task of Part III to assess the normative weight of reasons stemming from well-being. First, however, we need to address the worries Scanlon raises concerning the possibility of a theory of well-being.

## Chapter 7      The Impossibility of Theories of Well-Being

### 1. Introduction

In his chapter on well-being Scanlon argues against the possibility of ever finding a theory of well-being. Scanlon shows that, at a very high level of abstraction, there are what he calls three fixed points that any plausible theory of well-being would have to recognise. The list of these fixed points, however, does not amount to a *theory* of well-being. Their nature, in fact, is so different that it cannot be captured by a unified account. I begin by stating Scanlon's charge against the possibility of theories of well-being in more detail by formulating it as a five-point charge. Subsequently, I tackle each point by either criticising its fitness as a critical point or by showing the way in which our account would obviate the difficulty it poses (§2). At the same time I show how Scanlon's account of well-being can be dissolved into a reason-to-desire account (§3) and finally I provide an analysis of the notion of self-sacrifice and consider whether it is used consistently by Scanlon (§4).

### 2. The charge against the possibility of theories of well-being

According to Scanlon, there are three fixed points that should figure in any plausible theory of well-being:

- (1) Certain experiential states (such as various forms of satisfaction and enjoyment) contribute to well-being.
- (2) Well-being depends to a large extent on a person's degree of success in achieving his main ends in life, provided that these are worth pursuing or rational.
- (3) Many goods that contribute to a person's well-being depend on that person's aims but go beyond the good of success in achieving those aims (friendship,

other valuable personal relations, and the achievement of various forms of excellence, such as in art or science).<sup>1</sup>

According to Scanlon, however, a theory of well-being is different from a mere list of the ways in which a person's well-being can be contributed to. What would make this list a theory is the capacity to provide the following things:<sup>2</sup>

- (a) A more unified account of well-being on the basis of which one could see why the diverse things Scanlon has listed as contributing to well-being in fact do so.
- (b) A clearer account of the boundary of the concept—the line between contributions to one's well-being and things one has reason to pursue for other reasons.
- (c) A standard for making more exact comparisons of well-being—for deciding when, on balance, a person's well-being has been increased or decreased and by how much.

Having set these as the tasks that a theory of well-being should be able to handle, Scanlon concludes:

I doubt that we are likely to find a theory of well-being of this kind. It does not seem likely, for example, that we will find a general theory telling us how much weight to assign to the different elements of well-being I have listed: how much to enjoyment, how much to success in one's aims, and so on. I doubt that these questions have answers at this level of abstraction. Plausible answers would depend on the particular goals that a person has and on the circumstances in which he or she was placed. Perhaps a theory might tell us which goals to adopt, or at least which ones not to adopt. It does seem that there are answers to such questions, but I do not think that they are likely to be delivered by anything that could be called a general theory. Even if there were such a theory, moreover, it would need to be not just a theory of well-being, but a more general account of what is valuable and worthwhile.<sup>3</sup>

---

<sup>1</sup> See Scanlon, 1998, 113-126.

<sup>2</sup> See Scanlon, 1998, 125.

<sup>3</sup> Scanlon, 1998, 125.

Let us sum up the points made by Scanlon in this quote as two further charges against the possibility of theories of well-being:

(d) We are not likely to find a *general* theory of well-being.

In order to find answers to questions such as what weight is one to give to the various things that contribute to one's well-being, one cannot be operating at the level of abstraction at which theories are normally formulated. And finally:

(e) Even if there were such a general theory it would not be a theory of well-being but a theory of what is valuable in general.

Our strategy in the next section is quite simple. Firstly, we will take a close look at Scanlon's three fixed points and decide (1) whether all three deserve to be fixed points and (2) whether our own reason-to-desire theory of well-being cannot accommodate whatever fixed points we agree with Scanlon there should be. If it can accommodate them, then we will have answered at least point (a) above, i.e., we will have shown that there is a unified account of well-being on the basis of which one could see why the diverse things that contribute to well-being in fact do so. Subsequently, we will discuss points (d) and (c) in this order. A discussion of points (b) and (e) will be postponed until §4.

### 3. Scanlon's fixed points

#### 3.1. Are there three fixed points?

We will not spend a long time on Scanlon's first fixed point. That certain experiential states such as satisfaction and enjoyment are part of our well-being can be accepted quite easily. It does not take much experience and reflection to see that pleasure or, as Sidgwick would say, pleasurable consciousness, is widely desired and desirable. Scanlon, however, takes it that not all the things that contribute to our well-being need be pleasurable in this sense. In fact he rejects *mental statism* i.e., the thesis that "[s]omething contributes to

someone's well-being if, but only if, it affects the quality of one's experience."<sup>4</sup> This is an important point as the literature on well-being is often divided on the question of mental statism. On this issue, we are in full agreement with Scanlon. The rejection of mental statism is Scanlon's reason for introducing his second and most important fixed point. Well-being depends *to a large extent* on a person's degree of success in achieving his main ends in life, provided that these are worth pursuing.

What is important for our purposes is to notice here that Scanlon *must* take it that his first fixed point cannot be entirely captured by his second fixed point. This follows from his claim that there cannot be a unified theory of well-being, i.e., that neither one of his three fixed points is entirely reducible to any of the others. In this case, then, there must be experiential states that can contribute to a person's well-being without being among that person's adopted and rational aims. In order to see why Scanlon must think so we must look at his second fixed point in detail. First, though, a few words on the third fixed point.

According to Scanlon, many goods that contribute to a person's well-being depend on that person's aims but go beyond the good of success in achieving those aims. Examples of such goods are friendship, other valuable personal relations, and the achievement of various forms of excellence. Scanlon illustrates this thought as follows:

A misanthrope, who cares nothing for friends but to whom others are nonetheless devoted, may get some of the instrumental benefits of friendship, such as the help that friends provide, but not those benefits of friendship that involve standing in a special relation to others, since he does not stand in that relation to anyone. It is debatable whether the life of such a person would be better if these people genuinely cared about him than it would be if they treated him in exactly the same way out of other motives. Even if it does make a difference, however, it does not make as important a difference as it would in the case of a person who himself cared about friendship and regarded these people as friends. But even though the greater difference that the genuineness of friends makes in the latter case depends on the person's having a certain aim, this contribution to well-being is not plausibly accounted for simply by the idea of success in one of one's rational aims.<sup>5</sup>

---

<sup>4</sup> Scanlon, 113.

<sup>5</sup> Scanlon, 1998, 123-124.

From this example, we gather that friendship may deliver two kinds of goods: the instrumental goods that friendships normally yield and the good of standing in a special relation to others. The life of the person who cares about (i.e., who has an aim of) standing in that special relation, once his aim is achieved, will be better in the two ways just mentioned. Yet, some of the instrumental goods of friendship stand to be gained whether or not one has adopted friendship as an aim. Importantly, we must note here that Scanlon uses two closely linked expressions: earlier in the quotation he talks of “someone’s life being made better” while in the last sentence he talks of a “contribution to well-being”. Does Scanlon take these two expressions to mean the same?

We would normally take it that something that contributes to a person’s well-being would make that person’s life better and if a person’s life is better for this person then his well-being has been contributed to. That’s not so for Scanlon. He believes that the broadest way in which we can understand the expression *how well a person’s life goes* is what he calls the choiceworthiness of this person’s life. To ask how well a person’s life goes is not simply to ask *how well-off he or she is* in terms of his or her well-being. Choiceworthiness is not simply to be understood in terms of well-being but also in terms of *worthiness* i.e., the degree to which the quality of a life is particularly admirable or worthy of respect. In order to clarify this distinction Scanlon tells us that the life of a person who sacrifices his own well-being for the sake of others may be, for that reason, a particularly valuable one, and in order for this to be true there must be a sacrifice involved.<sup>6</sup> It follows on Scanlon’s account that the life of a person who is sacrificing his well-being in order to make a worthy choice, is choiceworthy, just like the life of a person who prefers a less worthy option but with a higher level of well-being.

We ought to ask, however, what Scanlon means when he writes that “someone’s life is made better” as in the last passage quoted above. Does he mean that the agent is better off, i.e., the agent’s well-being has been contributed to? Does he mean that the agent’s life is more worthy? Or does he mean both? Scanlon concludes the paragraph about the misanthrope as follows:

---

<sup>6</sup> Scanlon, 1998, 112.

The point is a general one: a life is made better by succeeding in one's projects and living up to the values one holds, provided these are worthwhile; but if these aims *are* worthwhile then succeeding in them will also make one's life better in other ways. This is true of friendship because standing in this relation to others is itself a good (albeit one that depends on one's having certain aims), and I believe the same can be said of, for example, the achievement of various forms of excellence.<sup>7</sup>

I believe Scanlon is here claiming that if the agent has adopted friendship as a rational aim then succeeding in this aim will contribute to his well-being. This is in accordance with Scanlon's second fixed point: well-being depends to a large extent on a person's degree of success in achieving his main ends in life, provided that these are worth pursuing or rational. But what does Scanlon mean when he says "but if these aims *are* worthwhile then succeeding in them will also make one's life better in other ways"? In light of his distinction between worthiness and well-being, it can only mean that the agent's life is made better with respect to its worthiness. Surely, however, if this is what Scanlon means (and I do not see what else he could mean), we do not have to consider this point as a third fixed-point in a theory of *well-being*. It might be an important point in a theory of what makes lives more worthwhile. But these would be two distinct kinds of theories.

As Scanlon himself puts it, a more worthy life may be a more admirable life, more worthy of respect. A theory of well-being, however, is concerned with the kind of value that life would have for the agent who lives it. The question here is whether an agent has reason to *desire* that life for himself. If he does and he is successful in his pursuit of this life we may say that a more worthy life will be a better life for the agent in terms of his well-being. But what makes a life more worthy or admirable is not necessarily what makes it more desirable for an agent. It may, as in the example we have just considered, or it may not.<sup>8</sup> In conclusion, then, there is no need for Scanlon's third fixed point in a theory of well-being though this point might appear as an element within a theory of what makes a life a worthy life.

---

<sup>7</sup> Scanlon, 1998, 124.

<sup>8</sup> We will have a chance to discuss this distinction later in this chapter as well as in Chapter 8. The discussion there concerns precisely the difference between two types of agents: those who have spontaneous rational desires for worthwhile aims and those who take themselves to have reason to pursue some worthwhile aims because they think it is their duty to do so.

### 3.2. Are there two fixed points?

In this subsection, we shall see whether Scanlon's second fixed point cannot accommodate things such as experiential states, i.e., the first fixed point. After all, why should it not be one of our rational aims to achieve or maintain a certain level of pleasurable consciousness in our lives? I guess Scanlon might agree. Yet, he may also claim that instances of pleasurable consciousness can make our lives better independently of our having adopted them as an aim. In other words, the first fixed point cannot be reduced to the second fixed point. Scanlon's notion of a rational aim, then, must be such as not to allow for this reduction. It may be worth pointing out that on the reason-to-desire theory, we need not distinguish between things that contribute to our well-being in such a way: we would simply say that agents have reason to desire pleasure just as achievement or special personal relations such as friendship, as long as these ends are indeed recognised as intelligible desirability features of agents' ends.

What we have to answer, then, is why Scanlon should think that there are instances of pleasurable consciousness that cannot be captured by his rational aim account of well-being. In order to answer that question we should follow the rationale that led Scanlon to choose rational aims as one of the fixed points, rather than, say, a reason-to-desire account. Scanlon begins his examination of theories of well-being by looking at actual desire satisfaction theories.<sup>9</sup> After dismissing this kind of theory, he turns to hypothetical desire accounts of well-being. According to him, these are affected by the following problem:

the idea of an informed desire is often understood as a purely hypothetical notion—what the person would desire under certain conditions—and is often used as a way of avoiding appeal to the normative idea of what a person has reason to desire. When “informed desire” is understood in this way, a notion of well-being based on it will lack a sufficiently close connection with what a person has a reason to want and to do. It may be likely that a person has reason to want those things for which he or she would have an informed desire, but this is by no means certain.<sup>10</sup>

---

<sup>9</sup> Scanlon, 1998, 113-114.

<sup>10</sup> Scanlon, 1998, 114-115.



Note that Scanlon seems to be aware of the difference between hypothetical desires and the “normative idea of what a person has reason to desire”. The point of this criticism is that the hypothetical notion of informed desires might lack a strong enough connection with a person’s well-being. I discussed this problem as it related to hypothetical accounts in the last section of Chapter 4 and I agree with Scanlon.

Scanlon thinks that one of the things that can be meant by saying that a person has a desire is that achieving or getting that thing is one of the person’s aims. Scanlon also notices that there may be both a wide and a narrow interpretation of the idea of rational aim. He wants to take on board the narrow one because:

... one cannot respond to every value or pursue every end that is worthwhile, and a central part of life for a rational creature lies in selecting those things that it will pursue. It thus makes a difference whether an aim has been adopted, and this is the rationale behind the narrow interpretation of “rational aim”...<sup>11</sup>

There might be a number of things that agents may rationally adopt as aims. But, of course, we cannot adopt all these aims. Scanlon thinks that when we adopt an aim we form an intention and the fact that an agent has adopted an intention makes a difference to the agent’s reasons:<sup>12</sup>

If I have merely decided (with good reason) that I intend to do a certain thing someday, but have not done anything about it, then the fulfilment of this aim does little or nothing to make my life better in the way we are presently discussing. The aims whose fulfilment makes a significant contribution to a person’s well-being are ones that that person has actually acted upon and, typically, given a role in shaping his or her activities and plans. The fulfilment of that aim, then, makes a difference to the person’s life by making these plans successful ones.<sup>13</sup>

Let us sum up the picture. There are so many things that we take ourselves to have equally good reason to adopt as aims. But we cannot pursue all of these things. Thus, we adopt some aims and not others. When I adopt an aim, I have the intention to pursue it. The fact

---

<sup>11</sup> Scanlon, 1998, 119.

<sup>12</sup> Though Scanlon thinks intentions are not independent sources of reason. Scanlon, 1998, 46.

<sup>13</sup> Scanlon, 1998, 121.

that I have gone at least some of the way towards fulfilling this aim changes my reasons. At this stage, it would no longer be as reasonable to choose the aims previously not adopted. This is Scanlon's narrow interpretation of rational aims. Now, typically, episodic and unplanned instances of pleasurable consciousness may make our lives better and may do so despite our failure to have adopted pleasure as an aim and having been actively engaged in fulfilling our aim. Their contribution to our well-being is independent of our having adopted them as rational aims. Hence, the rationale for Scanlon's separation between the first and the second fixed point.

In the following paragraphs I will attempt to show that (1) Scanlon's distinction between things that contribute to our well-being as part of a successful rational plan of ours and things that contribute to our well-being without being part of any rational plan of ours should be played down rather than emphasised; (2) if we can revert from 'aim' to 'desire' and don't make rational aims a subclass of aims, we get the reason-to-desire view.

Imagine someone—call him *Libero*—who abhors precisely having important aims structuring his life. *Libero* dislikes rational plans and working towards their achievement. He “goes with the flow” and is usually never in the same job, town, and relationship for more than a short period of time. He loves change and feels as a loss of freedom anything that prevents change from taking its course. I think a theory of well-being should allow that characters such as this one might enjoy a life that is good for them. That is because it seems to be perfectly intelligible that a person found desirable such a life and the goods that it offers, at least for a period of his or her life. Someone may retort that this person's rational plan is that of having no rational plan. Though that would sound paradoxical, I suppose it would be possible to say it. In his defence, Scanlon may also use the following thought:

The term 'aim' invites an interpretation that is ... voluntaristic...: an aim is something one “adopts,” and having an aim is a matter of intending to bring about a certain result. For present purposes, however, 'aim' needs to be understood in a way that is broader than its normal meaning.... If I have the aim of being a good son, then succeeding in this

contributes to my well-being even though there was no moment at which I “adopted” this aim or consciously formed this intention.<sup>14</sup>

Libero, for example, need not have consciously adopted the aim and the intention of “going with the flow” though this aim plays an important role in structuring his activities and plans.<sup>15</sup> But on this non-voluntaristic (and rather loose) meaning of ‘aim’ Scanlon’s distinction between the broader and the narrower interpretation of rational aims begins to lose its force. Remember that the rationale behind the narrow interpretation of “rational aim” is that we cannot pursue every valuable end and thus “it makes a difference whether an end has been adopted” and “actually acted upon.” But now it is hard to see how ‘having “adopted” aim  $x$ ’ would differ from ‘being disposed to  $x$ ’. “Adopting” the aim of going with the flow, for example, would amount to no more than having the disposition to see that aim as desirable for one. Similarly, it would now be possible to say enjoyment is an aim someone has “adopted” even if this person never adopted it or consciously formed the intention to pursue it. But now it really sounds as there were no difference between Scanlon’s first and second fixed point.

Scanlon may insist: “adopting” one end rather than an equally rational one and beginning to work towards its achievement changes the reasons we have. But the reason-to-desire account can capture this thought without any appeal to the “adoption” of aims. On our account, one may say that at  $t$  there is reason of equally strong degree to desire to pursue two courses of actions, but that at  $t+1$ , when one of the two courses of action has been chosen and pursued to some extent, there is stronger reason to desire and pursue the chosen course of action. On this account, the difference is dictated by the degree to which there is reason for this person to desire both courses of action at a time; at each time, this degree is determined by facts about the agent, including his circumstances and dispositions.

As a final defence, Scanlon may want to press another charge against the reason-to-desire account:

---

<sup>14</sup> Scanlon, 1998, 119-120.

<sup>15</sup> Note that someone wanting to defend the explanatory importance of the concept of well-being *contra* Scanlon, may say something similar of the aim of pursuing one’s well-being: we need not have consciously adopted it though it structures our plans and activities.

Even if we ... identify “informed desire” with “rational desire,” there remains the problem that the objects of a person’s informed desires are likely to include many things that are not related to the quality of the desirer’s own life, intuitively understood. Suppose for example that I very much admire a certain person, and therefore desire that her struggle and sacrifice will be crowned with success and happiness. This may be a rational desire as well as an informed one; it might, quite properly, be strengthened by fuller knowledge of the person’s life and character. Even if this is so, however, if I have no connection with her beyond my admiration and this desire, then the quality of my life is not affected one way or another by her fate.<sup>16</sup>

We have already examined this type of charge in Chapter 5, as it was put forward by Darwall. Given the threat this kind of charge poses to our account, however, we will tailor a reply to Scanlon’s own version of it.

Matthew has reason to desire that Jane, a person he very much admires, will find success and happiness in her struggle. And yet, if Matthew has no connection with Jane beyond his admiration and this desire, then the quality of his life, intuitively understood, is not affected one way or another by her fate. Why think that the attaining of the object of Matthew’s rational desire will not contribute to his well-being after all? If Matthew has a spontaneous genuine desire that Jane achieves success in her admirable ends, he will conceive of her success as part of what makes his life go better. To at least some extent, he identifies his well-being with Jane’s success. This is the familiar phenomenon we described as the expansion of the self. However, if this description was not true of Matthew, then we should refrain from claiming that he has a genuine spontaneous desire that Jane achieves success in her aims. The ‘desire’ appealed to by Scanlon might be no more than a whim. The fact that Matthew’s desire might be strengthened by fuller knowledge of Jane’s life and character, however, indicates that Matthew’s is a case of the expansion of the self.

If the preceding considerations are correct, there is no reason to emphasise the importance of rational aims in the first place, as in fact Scanlon’s own “narrow interpretation of rational aim” seems to be so loose as to dissolve into a reason-to-desire account. In conclusion, then, and *contra* Scanlon’s criticism (a), there is no reason to believe that

---

<sup>16</sup> Scanlon, 1998, 115.

Scanlon's three-fixed points account could not be reduced to a unified theory of well-being such as ours. On this latter view, just as we may have reason to desire particular non-aimed-at instances of pleasure (or other goods) we may have reason to desire, say, that our lives never fall below a certain threshold level of enjoyment.

### 3.3. The generality of theories and well-being comparisons

At this point, Scanlon might retort that our account of well-being is not a *theory* because it is not still clear that it is a *general* enough account of the kind theories should be (this is his point (d) in §2 above). To define well-being as what there is reason for an agent to desire is, I believe, as general and abstract an account as any theory should be. Of course, to say that well-being is defined in this way is not yet enough to say that we know for each and every agent what is most in her interest at any time. But no theory ever needs to be like that, not even in the realm of the hard sciences. When we say that  $e = mc^2$  and we want to calculate the energy of a given body we still have to know the value of its specific mass. Similarly, we can abstractly say that the well-being is what there is reason for an agent to desire but if we need to know what the well-being of a particular agent is, we need to provide some special facts about the agent.

We are now left with Scanlon's criticism (c), namely, that our theory of well-being cannot be considered as a theory, for it cannot "provide a standard for making more exact comparisons of well-being—for deciding when, on balance, a person's well-being has been increased or decreased and by how much." The question of comparative judgements of well-being is a difficult question as applied to both intrapersonal and interpersonal cases. The literature on this topic is vast and it is not my intention here to enter this discussion in great detail. I shall, however, make a few remarks concerning the way in which I believe our account of well-being should tackle this issue.

When looking at comparisons of well-being—both intrapersonal and interpersonal—we are looking at what measure (quantity and quality) of well-being an agent stands to gain or lose if a given state of affairs obtained. The measure of well-being of a person is cashed out in terms of the arrangement of the various parts constituting the agent's well-being. At this stage, our theory of well-being is of help in individuating what kinds of things are

good for agents: pleasure, knowledge of one's situation, etc. What arrangement would be best for each agent, however, can only be determined when considering particular facts about the agent in question and her circumstances. Even here, however, our theory of well-being can be helpful. On our account, *ex ante* and *ex post* desires are taken to be the primitive criteria for well-being claims. We may then begin to look at the agents' respective desires to begin to assess the impact of certain states of affairs on their respective well-being.<sup>17</sup>

I am not here claiming that on our theory it is possible to make exact comparisons of well-being. For one thing, however, our theory may help us understand why our capacity to make such exact comparisons is limited. We may also wonder whether Scanlon really takes it as a necessary requirement for any theory of well-being to be able to provide exact comparisons. That may be inappropriate: Aristotle remarked long ago that our account of politics, and we could also add ethics, will be adequate if it achieves such clarity as the subject-matter allows; for the same degree of precision is not to be expected in all discussions. Even if Scanlon agreed with Aristotle, however, he may still want to undermine the plausibility of our theory *qua* theory by appealing to points (b) and (e) above. We shall address these criticisms in the next section.

#### 4. Scanlon, choiceworthiness, and self-sacrifice

Consider (e): no such thing as *just* a theory of well-being can be found, as we can at most devise a more general account of what is valuable and worthwhile. This claim is connected to (b): we are not likely to find a theory of well-being capable of providing "a clearer account of the boundary of the concept of well-being—the line between contributions to one's well-being and things one has reason to pursue for other reasons." Scanlon's argument supporting (e) and (b) is simple enough:

---

<sup>17</sup> Of course one needs to assess whether the agent's spontaneous desires are genuine and converge with the judgement of competent judges.

Succeeding in one's main aims, insofar as these are rational, must be a component in any plausible notion of well-being. But this idea serves as an evaluative Trojan horse, bringing within the notion of well-being values that are not grounded in it.<sup>18</sup>

We adopt many rational aims; some of them an agent has reason to pursue insofar as they contribute to her well-being, others insofar as they are worthy aims grounded on values other than well-being. However, (via fixed point two) whenever we achieve success in any of these rational aims, even those not grounded in our well-being, our well-being is thereby contributed to. Hence, we cannot provide an account capable of drawing the line between contributions to one's well-being and things one has reason to pursue for other reasons (cf. (b)) and, thus, we cannot find a theory of well-being but, at most, a more general account of what is valuable or worthwhile (cf. (e)).

In Chapter 3, I already offered a reply to this argument: it is not the case that all worthwhile pursuits we have reason to will, will contribute to our well-being. Only those pursuits we have reason to desire will do so. We shall not rehearse this argument here and will take it that Scanlon's criticisms (b) and (e) have safely been addressed. Also, in Chapter 5, I discussed the way our own account deals with the scope problem. An account capable of preserving the boundaries between well-being and rational self-sacrifice is of course capable of drawing the line between contributions to one's well-being and things one has reason to pursue for other reasons. Scanlon, however, makes repeated appeals to the notion of self-sacrifice though, as we shall see next, it is not clear how his own account is to deal with the scope problem.

#### 4.1. Choiceworthiness, worthiness, well-being, and sacrifice

First, we shall remember that on Scanlon's account the question of "how well a life goes for a person" from that person's point of view is understood as how "choiceworthy" that life is for the agent. Choiceworthiness is in turn understood as well-being and worthiness. Most importantly, Scanlon appeals to the notion of self-sacrifice to draw a line between pursuits that would make a person's life more worthy and pursuits that would contribute to that person's well-being:

---

<sup>18</sup> Scanlon, 1998, 133.

The life of a person who sacrifices his own well-being for the sake of others may be, for that reason, a particularly valuable one, and in order for this to be true there must be a sacrifice involved. ... We might say, for example, that there is a reason to choose a certain life because of its great value, even though it involves a low level of well-being, or that the value of a life did not in fact make it worth choosing given the sacrifice in well-being that it would involve.<sup>19</sup>

I agree with Scanlon insofar as he claims that there may be reason for an agent to *choose* or *pursue* a certain life because of its great value even though it involves a low level of well-being. The question that one should ask, however, is whether there is reason for an agent to *desire* that life for himself. If there were, then pursuing that life would be in the agent's interest; one would live both the life that is best for one and the most worthy life. The virtuous agents described in the last chapter may be the kind of agents that would understand worthy lives in these terms. Even if an agent took himself to have no reason to *desire* such life, however, he might nonetheless take himself to have reason to live that life. What is unclear is why Scanlon would claim that choosing this life *in this case* would make the agent's life "go better" from the agent's point of view. Contrary to what Scanlon thinks a more choiceworthy life cannot simply be equated to a "a life that goes better for the agent." That's even more so when one wants to say, as Scanlon does, that this life would involve a sacrifice.

To sum up this first criticism we can say that Scanlon has wrongly defined the idea of a life's choiceworthiness in terms of "how well a life goes for the agent". We will return to this point later on. Our second, third, and fourth criticisms are also linked to this point. In order to discuss them we shall introduce an example Scanlon provides in a footnote:

It may seem that when we [distinguish worthiness from well-being] we are identifying well-being with experiential quality, and that when those two are carefully distinguished the question of well-being turns out to be the same as the question of choiceworthiness. But this is not so. A person who abandons a valued ambition in order to help his family may have made a net sacrifice in the quality of his life, by giving up the accomplishments he would have made, even if the experiential quality of the life he chooses is no lower than

---

<sup>19</sup> Scanlon, 1998, 112.



that of the life he forgoes. It may, for example, involve more joy and less struggle, stress, and frustration. The life he lives could therefore be more choiceworthy and involve no loss in experiential quality while still being a worse life for him, in the sense in which I am concerned here.<sup>20</sup>

Let us elaborate this example. Jones is facing the following choice between two possible rational aims: either pursue his career as an architect but abandon his family, or live with his family whilst renouncing his career ambitions. Suppose now that Jones decides to stay with his family. Scanlon wants to say that this option involves Jones' *sacrifice* of his career ambitions. Scanlon wants us to believe that Jones goes for the worthy choice against his well-being. Jones has to choose between two rational aims and, if he chooses rationally, he will choose that which he has most reason to choose. But now, and this is my second criticism, on the basis of what can Scanlon claim that the option Jones has most reason to choose involves a greater sacrifice on Jones's behalf? What is to draw the line between pursuits involving sacrifices and pursuits that don't? On Scanlon's account it may very well be the case that choosing his career over his family would constitute an even greater sacrifice for Jones. This account simply lacks the conceptual means to even get started as to what contributes to one's well-being and what doesn't.

My third criticism has to do with the second fixed point. Success in any of a person's rational aims is sufficient to contribute to this person's well-being. Suppose that Jones for some reason does not choose the option that is best in terms of his well-being. Even though his well-being *would not be maximised* by his choice, given the second fixed point, we can assume that his well-being *would nonetheless be contributed to*. But is it always the case that success in any of our rational aims contributes if only even minimally to our well-being? It is certainly not clear from the phenomenology of our ethical life. At times agents do act with the belief that the action that they feel compelled to do would not benefit them in any way. Is this belief always a wrong belief? If it were, then it would have to be dispelled as an error. But Scanlon offers no such error theory.

---

<sup>20</sup> Scanlon, 1998, 385-386, n. 4. Scanlon asks us to assume that the agent's experiential quality would be equal in the two cases. This example is meant to show that how well someone's life goes cannot be accounted by well-being alone, as well as that there is a sense in which well-being is broader than experiential quality. We agree with Scanlon's latter claim and will henceforth ignore it.

Fourthly, Scanlon's decision to define choiceworthiness as "how well a person's life goes for him" blurs the notion of rational self-sacrifice. If achieving success in one's rational aims always increases one's life choiceworthiness, can we say, as Scanlon would want us to, that someone like Jones makes a net sacrifice? Of course we may suppose that he chooses one of the two options that involves the least well-being for him. And yet, on Scanlon's own terminology, if the agent acts rationally, we would still have to say that this option is the one that makes his life most choiceworthy and, hence, the best life for him. Suppose Jones chooses to stay with his family and that makes him miserable (great loss in his well-being). Yet, this is what he takes himself to have most reason to do. Would Jones say (from a first-personal, non-moral point of view) that his life is on the whole better for him?

These points tell against both Scanlon's rational aim account of well-being as well as his charges (b) and (e) against a theory of well-being. A theory of well-being is needed precisely because we take it that there is an important boundary between things that contribute to one's well-being and things that involve a rational self-sacrifice. This latter notion is also blurred by Scanlon's conception of choiceworthiness: how well the life goes for a person and how good (worthy) the life of a person is, are two clearly distinct things. Though an agent may take himself to have reason to pursue a good (worthy) life, it is nowhere clear that a more worthy life will be better for him in any sense, hence it is not clear that he has reason to *desire* it. In the next two subsections we shall consider the notion of sacrifice further.

#### 4.2. Sacrifice and spheres of compensation

Scanlon claims that it would be wrong to think that well-being may constitute a significant category of goods on the basis that it constitutes a distinct "sphere of compensation." It is important to look at this argument here because in it Scanlon makes frequent appeals to the notion of sacrifice. For well-being to be considered as belonging to a sphere of compensation of its own it would have to be true that losses in well-being of one kind at one time could be fully made up for by other gains in well-being, but not by considerations of other kinds:

Even if other considerations constitute good reason for accepting a loss in well-being, this loss remains a loss, but (the suggestion runs) when we give up one element of well-being for another (such as when we give up a pleasure now for the sake of a greater pleasure later) there is no real loss. This might be put by saying that well-being constitutes a distinct “sphere of compensation.” This idea is appealing, but mistaken. We do speak of making a sacrifice when, for example, we give up comfort and leisure for the sake of a family member or a friend, or for the good of some group, team, or institution of which we are a member. But it also feels like a sacrifice when we give up present comfort and leisure for the sake of our own longer life or future health. The fact that in the latter case we will be “paid back” in the same coin, our own well-being, does not make this case less like a sacrifice than the other at the time that it is made. The term sacrifice is appropriate in both cases because we give up something of present, palpable appeal for the sake of some other, possibly more distant concern.<sup>21</sup>

The thought is that it certainly feels as if we are making a sacrifice both in the case in which we are giving up present well-being for the sake of things other than well-being as well as in the case in which we are giving up present well-being for the sake of future well-being. The term sacrifice, Scanlon claims, is appropriate in both cases. Hence, the notion of sacrifice is not sufficient to single out any special category of goods constituting well-being. There are several remarks to be made about this argument.

Firstly, we need to note that this passage does indeed suggest that Scanlon takes the notion of sacrifice not to be peculiar to any one sphere of our lives but to our rational action in general. The notion of sacrifice is defined intuitively as “the loss of something of present or palpable appeal for the sake of some other, possibly more distant concern.” Now take the case of Jones again. On this definition of sacrifice it seems very plausible to say that either choice involves a sacrifice. Even if Jones chose the option involving the greater sacrifice that would not guarantee that he chose the option that leaves him with the least well-being. The notion of sacrifice then seems completely disconnected from that of well-being.

In connection to this point, we shall soon discuss the work of Raz as someone who *clearly* states that well-being and sacrifice are not essentially connected. Scanlon’s account,

---

<sup>21</sup> Scanlon, 1998, 127.

however, is incoherent when it claims both (a) that the notion of sacrifice cannot drive a wedge between well-being and other kinds of goods and (b) that the “life of a person who sacrifices his own well-being for the sake of others may be, for that reason, a particularly valuable one, and in order for this to be true there must be a sacrifice involved.” If (a) is true, the notion of sacrifice cannot distinguish losses of well-being from other kinds of losses. If that is correct then the fact that someone is sacrificing does not show that his choice is less than optimal in terms of his well-being. Hence we cannot assert that his life is particularly worthy or valuable because it contains a sacrifice in well-being.

Secondly, Scanlon compares the loss of well-being that one may undergo for the sake of a friend or a family member with the loss of one’s well-being for the sake of future well-being. But are these really differing cases? Is it not rather the case that in the first case one identifies one’s well-being with one’s friend’s or family member’s, and even with the interests of the institutions to which one belongs? If one does understand these cases in this way then there would be no difference between them and the case of one’s loss of present well-being for one’s future well-being.

Thirdly, something ought to be said about Scanlon’s notion of sacrifice. Our going to the dentist certainly *feels* like a sacrifice even if in the long run it might be advantageous in terms of well-being. If that is the case, then, having the *feeling* of sacrifice is not sufficient to distinguish a present loss of well-being for a future advantage, as opposed to a present loss of well-being for no future advantage. Whatever our *feelings* might be, however, most of us would think that there is a great difference between, say, (a) undertaking a great labour for a great reward; and (b) undertaking a great labour for no reward. While I can see why we would have reason to desire (a) for its sake, I can’t see why we would have reason to desire (b) for its sake. My guess is that this is so precisely because while it is not clear that (a) involves a sacrifice of our well-being, (b) suggests precisely that. I think the notion of sacrifice and its link to the notion of well-being need to be given further attention.

#### 4.3. Sacrifice, self-sacrifice and well-being

Raz writes:

...while it seems clear that people sometimes do and sometimes have to make sacrifices for moral reasons, I do not believe that we always harm our self-interest, or act against our expectations regarding our self interest, when we make sacrifices. Once we see that self-sacrifice is not essentially connected to well-being, it will be easier to understand why an agent's well-being is not, normally, a reason for action for that agent.<sup>22</sup>

Consider the last sentence, and focus only on the idea that well-being is not essentially connected to self-sacrifice. Let us also note the shift between talk of 'sacrifice' in the first sentence and talk of 'self-sacrifice' in the second. Next, let us look at Raz's characterisation of sacrifice:

We mark types, or degrees of sacrifice. Whenever we give up (for what appears a good reason) something that we deeply care about, we are making a sacrifice. Normally this would warrant saying that we are sacrificing the goal or relationship we are giving up (e.g., our career). When we do so with great reluctance, we also sacrifice ourselves or (in slightly different circumstances) our life.<sup>23</sup>

In what follows I will disagree with Raz's claim that well-being and self-sacrifice are not essentially connected. Take our example of Jones. Suppose that both aims would contribute to Jones's well-being and Jones chooses the one that contributes most to his well-being. In this case, I would claim there is no *self-sacrifice*, that is, Jones is not *sacrificing his well-being*; the only case in which Jones's rational self-sacrifice would occur is the case in which he has most reason to do (and does) precisely that which is not the best available option for him. I take this thought to be expressing the idea that well-being and self-sacrifice are essentially connected.<sup>24</sup>

Raz may retort that, surely, when Jones chooses his family over his career he is sacrificing something he deeply cares about, namely his career, even if a family life is what is best for him. I would want to concede this point to Raz. It is reasonable to say that Jones is *sacrificing his career* to his family life. Thus, we could agree with Raz when he says "I do not believe that we always harm our self-interest, or act against our expectations regarding our self interest, when we make sacrifices." *Contra* Raz, however, I think we must

---

<sup>22</sup> Raz, 2000, 221.

<sup>23</sup> Raz, 2000, 223, n.18.

distinguish the notion of “sacrificing something we care about for something else we care (even more) about” from the notion of a *self*-sacrifice. One can sacrifice one’s career and still choose the option that his best for one: no self-sacrifice is involved in this case. Jones, for example, may simply realise that he does not live in a world in which his career and his family life are compatible pursuits. Tough luck, he may think, but which one of these two pursuits do I *want* most? If he chooses his family life, there is a sense in which he sacrificed his career. But this sense would be very different from the case in which Jones chose what he clearly saw as being the less good option in terms of his own well-being. That choice would be self-sacrificial.

Going back to Scanlon with this analysis of the notions of sacrifice and self-sacrifice we can say the following. Self-sacrifice only occurs when my life’s choiceworthiness is increased at the expense of my well-being. Sacrifices, however, may occur also when our well-being is nonetheless maximised by our choices.<sup>25</sup>

## 5. Conclusion

Taking Scanlon’s fivefold charge against the possibility of well-being we can now claim the following. Our reason-to-desire theory of well-being provides a unified account on the basis of which one can say why the diverse things contributing to well-being do in fact do so. It affords a clearer account of the boundary of the concept—the line between contributions to one’s well-being and things one has reason to pursue for other reasons. Of all the things we have reason to pursue, only those we have reason to desire will contribute to our well-being. In this chapter I showed that once certain criticisms of Scanlon’s

---

<sup>24</sup> We can also think of self-sacrifice as of *net sacrifice*.

<sup>25</sup> This distinction may also help to explain another phenomenon of our ethical life. Imagine two characters Vincent and Deborah always ready to sacrifice their free time in order to keep the department of history at which they both work functioning at its best. Vincent, however, does this virtuously; he identifies his good with that of the department. Deborah, on the other hand, does this dutifully. She thinks she has reason to do this but she does not see it as a good thing for her at all. On one occasion, the department contacts both Vincent and Deborah at their respective holiday houses: their help is badly needed. Both Vincent and Deborah decide to shorten their holidays to come and help, though one does so virtuously and the other dutifully. On their arrival the head of department thanks them both and, we would think, has good reason to

account of well-being are taken into consideration, his three fixed points can be expressed more simply as what there is reason for an agent to desire. I also argued that it might be simply too much to expect an ethical theory to provide “more exact comparisons of well-being” and to be more general than the one I have offered. Finally, I have shown the problems involved in defining choiceworthiness as “how well a life goes” for the person whose life it is. I claimed that one cannot use this definition of choiceworthiness and at the same time claim, as Scanlon does, that sacrifice (at least when intended as self-sacrifice) drives a wedge between goods that contribute to the worthiness of our lives and goods that contribute to our well-being.

---

do so. Insofar as they both *sacrificed* something for the department they deserve to be thanked. That is true even for Vincent, though we cannot be sure that Vincent’s action was *self-sacrificial*.

## Part III Moderate Welfarism

### Chapter 8 Welfarism and Moderate Welfarism

#### 1. Introduction

In this chapter, I will examine welfarism as defined by Scanlon: *well-being is the only value or the "master value", all other goods ultimately deriving their value from well-being*. In fact, I will refer to this view as to *radical welfarism*. I will offer two reasons to reject this view and on my way to doing this I will defend the buck-passing account of goodness from one important challenge (§2) and clarify the relation between our reasons to desire things and our practical reasons (§3). With Scanlon, I will agree that there are kinds of values that are irreducible to well-being. It follows that not all valuable things give us reason to desire them. In this chapter, however, I will put forward a more moderate form of welfarism. On this view, when no one else is affected by the agent's choices and actions, there can never be most reason for an agent to do something that is not what he has most reason to desire, something that doesn't bring about the best outcome in terms of his well-being (§4). This view is a welfarist view insofar as the agent's own well-being would be the ultimate criterion of practical reason.

#### 2. The buck-passing account: a first challenge to radical welfarism

##### 2.1. The buck-passing account of goodness and radical welfarism

We have already made frequent appeals to the buck-passing account of goodness throughout the thesis and showed how our own account of well-being is a specification of



such an account. In what follows, I shall offer a more detailed presentation of this account, I shall show how endorsing this account counts against radical welfarism, and I shall defend this account from one recent challenge. Scanlon writes:

[Contrary to Moore, I believe that] being good, or valuable is not a property that itself provides a reason to respond to a thing in certain ways. Rather, to be good or valuable is to have other properties that constitute such reasons. Since a claim that some property constitutes a reason is a normative claim, this account also [i.e., like Moore's] takes goodness and value to be non-natural properties, namely the purely formal, higher-order properties of having lower-order properties that provide reasons of the relevant kind. ... it is not goodness or value itself that provides reasons but rather other properties that do so. For this reason I call it a buck-passing account.<sup>1</sup>

To be good or valuable would amount to having other properties constituting reasons to respond to a thing in certain ways. More specifically:

To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it. Exactly what these reasons are, and what actions and attitudes they support will be different in different cases. ... To claim that something is *valuable* (or that it is "of value"), is to claim that others also have reason to value it, as you do.<sup>2</sup>

It might also be noted here that Scanlon is a pluralist concerning value-bearers. He thinks that we value many different kinds of things, including objects and their properties (such as beauty), persons, skills, and talents, states of character, actions, accomplishments, activities and pursuits, relationships, and ideals.

Now consider radical welfarism:

(RW) Well-being is the only ultimate good. All goods ultimately derive their value from well-being.

---

<sup>1</sup> Scanlon, 1998, 97.

<sup>2</sup> Scanlon, 1998, 97.

Next consider the fact that the value of some *accomplishments* is judged, partly at least, by its product or *outcome* or in the light of the *performance* of their constitutive acts. As Roger Crisp points out:

Outcome-assessments may, for example, make mention of aesthetic values. To understand Michelangelo's accomplishment properly requires understanding the unity, energy, and impressive clarity of the Sistine ceiling. Likewise, distinguishing Bohr from John Rawls's 'grass counter', whose aims may include counting blades of grass on various lawns, involves appeal to a notion of importance or significance which is not to be cashed out in welfarist terms. Performance assessments, similarly, involve various excellences, the value of which cannot be reduced to the value of well-being. Ruth's controlled exercise of his judgement and of his physical strength partly constituted his accomplishment.<sup>3</sup>

Even on a particularly wide understanding of well-being, values such as the aesthetic value of an accomplishment cannot be reduced to well-being. It is not on the basis of what is good for an agent that we appreciate whether an object is aesthetically more worthy than another. This, however, would have to be what we should say as radical welfarists. That this position is fundamentally wrong is shown even more patently by normative expressions such as 'good evidence'. If radical welfarism were true, good evidence must mean 'prudentially good for some being or type of being'. But can anyone plausibly hold the position that what qualifies some piece of evidence as good, is the fact that it is good for someone in this sense? This claim is implausible.

If radical welfarism were true, we would have reason to *desire* aesthetic beauty, good evidence, and every other valuable property. But do we reasonably respond to a beautiful painting by desiring it? That does not make sense. On the contrary letting go of radical welfarism but still embracing the buck-passing account of value we may say the following: an object *o*, or state of affairs *s*, is good if, and only if, it displays other properties that provide reason to have a certain pro-attitude towards it, i.e., that give one reason to choose it, admire it, prefer it, recommend it, etc. For example, *o* is good evidence that *p*, if it gives us reason to believe that *p*. Or again, suppose that a certain object is of aesthetic value. In this case, the buck-passing view may say that the object instantiates properties that give one reason to *admire* it. In conclusion, we should say that Scanlon is right in thinking that

there are different kinds of values, that well-being is not a master value and thus that radical welfarism is false.

## 2.2. A challenge to the buck-passing account

In ‘The Strike of the Demon’ Wlodek Rabinowicz and Toni Ronnøw-Rasmussen level a criticism to accounts of value of the kind defended by Scanlon, which they generically refer to as ‘fitting pro-attitude’ accounts of value.<sup>4</sup> Their charge is quite simply expressed:

It seems that the analysis in question [the fitting pro-attitude analysis] at best gives a necessary but not sufficient condition for value. That there are reasons to have a pro-attitude toward an object does not entail that the object in question must be valuable. ... the reason for a pro-attitude towards an object may have to do with the value of that attitude itself. Such a reason may exist even if the object of the attitude is devoid of value.<sup>5</sup>

Their charge is inspired by an example given earlier by Crisp:

since there can be reasons for preferring things that have nothing to do with their value, definitions [of value] in terms of preferability are false. Imagine that an evil demon will inflict a severe pain on me unless I prefer this saucer of mud; that makes the saucer well worth preferring. But it would not be plausible to claim that the saucer’s of mud’s existence is, in itself, valuable.<sup>6</sup>

Crisp’s argument is a *reductio ad absurdum* of the claim: ‘if there is reason to prefer *x*, then *x* is valuable’. In order to assess this argument we need to determine exactly what is meant by ‘there is reason to *prefer* a saucer of mud’. Crisp – after Kupperman – interprets it as meaning that we have reason to prefer its existence to its non-existence. However, it might also mean that we have reason to prefer our eating it to our not-eating it, our admiring its texture to our not admiring its texture, our recommending it to a friend to our not recommending it to a friend, etc. For the sake of simplicity, I will take on board

---

<sup>3</sup> Crisp, 2000b, 266.

<sup>4</sup> Rabinowicz & Ronnøw-Rasmussen, unpublished.

<sup>5</sup> Rabinowicz & Ronnøw-Rasmussen, unpublished, 9.

<sup>6</sup> Crisp, 2000a, 459.

Rabinowicz and Ronn w-Rasmussen's resolution to substitute the attitude of 'favouring' as generically representing all pro-attitudes.<sup>7</sup> Crisp's example is meant to show that we would have reason to favour the saucer of mud's existence. Given that, unlike Kupperman, we are not operating under the assumption that states of affairs are the only bearers of value, we can simplify further by claiming that the example shows that we would have reason to *admire* the *saucer of mud* (rather than *favouring its existence*). However, according to the demon example, we have reason to admire it not in virtue of any admirable property inherent to the saucer of mud. The example as used by Rabinowicz and Ronn w-Rasmussen is intended to show that *value of no kind, final or instrumental, supervenes on the saucer of mud and yet we have (instrumental) reason to admire it*.

### 2.3. In defence of the buck-passing account

What is the object of the agent's *valuing* in this example? Rabinowicz and Ronn w-Rasmussen would say it is the saucer of mud. The defender of the fitting pro-attitude account, however, defines *value* in terms of *there being reason to favour* something; he would then put this question as follows: 'what object, state of affairs, person, attitude, etc., is there reason to favour'? The answer she *should* and *could* give in this case is: 'the agent's *admiring the saucer of mud*'. Notice the shift from 'there being reason to admire *the saucer of mud*' to 'there being reason to favour *admiring the saucer of mud*'. The buck passer *could* think this because nothing either in theory or in the example in question would stop her from thinking that one should favour one's admiring of the saucer of mud as that would avoid the pain of being stricken by the demon. She *should* think so if she is to abide by the requirements of her own account of value. According to this account in order for an agent to have *reason* to favour an object (where that could be a thing, a state of affairs, a person, an attitude, etc.) there must be a property in the object that provides reason to have a pro-attitude towards it. Given that no such property ever supervenes on the saucer of mud, that cannot be the object of one's rational favouring. Only the attitude of admiring the saucer of mud can be taken (as a whole) as the object of one's rational

---

<sup>7</sup> The attitude of favouring might not be quite appropriate. Take the following example. A piece of evidence is good if and only if there is reason to favour it. But what would it mean 'to favour' a piece of evidence? Clearly, in this case, believing would be a more appropriate response. In what follows, I shall leave this problem aside and use 'favouring' flexibly.

favouring, as pleasure (or the avoidance of pain that would ensue from admiring the saucer of mud) does supervene on *it*.

Is this enough to stop the demon example from counting as a counterexample to the buck-passing account of goodness? As long as we can claim that there are instances in which having a reason to have a pro-attitude does not entail value, the buck-passing account cannot be maintained. The fact that the buck-passer can give a coherent explanation of the demon example does not stop the demon example from counting as a counterexample to her theory. She must also be able to show that it cannot possibly count as such, that there is some inherent problem with it. Schematically, the counterexample has this form:

If *F-ing* is the avoidance of my pain, then there is reason for me to *F*.

In this case, *F-ing* is admiring the saucer of mud.

Therefore, there is reason for me to admire the saucer of mud.

The fact that there also is reason to favour admiring the saucer of mud would not dispel the force of the counterexample. However, the buck-passer might advance an argument to the effect that there is something implausible with this counterexample. Suppose that on another occasion *F-ing*, i.e., what would make me happy, was my believing in God rather than admiring the saucer of mud. Does it follow that there is reason to believe in God? I suppose many would disagree. This point can be made rather emphatically through an example crafted for different purposes by the English mathematician and philosopher William Clifford (1845-1879):

A shipowner was about to send to sea an emigrant-ship. He knew that she was old, and not over-well built at the first; that she had seen many seas and climes, and often had needed repairs. Doubts had been suggested to him that possibly she was not seaworthy. These doubts preyed upon his mind, and made him unhappy; he thought that perhaps he ought to have her thoroughly overhauled and refitted, even though this should put him to great expense. Before the ship sailed, however, he succeeded in overcoming these melancholy reflections. He said to himself that she had gone through safely so many voyages and weathered so many storms that it was idle to suppose that she would not come safely home from this trip also. He would put his trust in Providence, which could hardly fail to protect these unhappy families that were leaving their fatherland to seek for better times elsewhere. He would dismiss from his mind all ungenerous suspicions about the honesty of builders

and contractors. In such ways he acquired a sincere and comfortable conviction that his vessel was thoroughly safe and seaworthy; he watched her departure with a light heart, and benevolent wishes for the success of the exiles in their strange new home that was to be; and he got his insurance-money when she went down in mid-ocean and told no tale.<sup>8</sup>

Let us refer to the belief that the ship would not sink as the belief that  $p$ . Did the ship owner have sufficient reason to believe  $p$ ? I hope everyone would agree in saying that he didn't. This conclusion relies on the following principle:

(E) There is reason to believe an empirical claim  $s$  only on the ground of features that make it credible, i.e., on the ground of the evidence in favour of  $s$ .<sup>9</sup>

That's because beliefs aim at truth and the truth of empirical claims is established through the available evidence in their favour. The fact that believing  $p$  was prudentially valuable for the ship owner, i.e., the fact that believing  $p$  would make him happy, light hearted and save him money, does not count as evidence in favour of  $p$ . Therefore, it does not create any reason to believe  $p$ . That, however, does not exclude that this fact could count as a reason to *bring it about* that the ship owner believed  $p$ . Notice the shift between having a reason to *believe*  $p$ , and having a reason to *bring it about* that one believes  $p$ . Questioning this conclusion would entail questioning (E). Similarly for the demon example. The fact that the demon would strike me, might at most be a reason to bring it about that I admire the saucer of mud, or as we put it above, it might give me a reason to favour admiring the saucer of mud. This example would of course rely on a principle different though parallel to (E):

(A) There is reason to admire an object  $x$  only on the ground of the features that make it admirable.

That someone would strike me unless I admired  $x$  does not make  $x$  admirable. This fact, however, might be good reason to *bring it about* that I admire  $x$ . My argument then has the following structure. To accept the demon example to count as a counterexample to the

---

<sup>8</sup> Clifford, 1879, vol. II, 177-178.

<sup>9</sup> Note that evidence may be taken to play a different role in justifying belief in normative claims such as (E).

buck-passing account of value involves rejecting a principle (A) analogous to a principle (E), which everyone should accept.

### 3. From affective to practical reason: a second challenge to radical welfarism

I defined a person's well-being as what there is reason for her to desire and put forward a list of the parts of well-being. However, saying that well-being is what an agent has reason to *desire* amounts to saying that our well-being reasons are *affective* reasons. Desires and the normative impulses that typically accompany them surely are affective responses. Yet, isn't it common to identify reasons having to do with one's well-being with *practical* reasons? We need then to flesh out the connection between affective and practical reason. As we shall see, doing so will bring out a second challenge to radical welfarism.

The connection between affective and practical reason is expressed by the following bridge principle, which we shall call FD or feeling/disposition principle:<sup>10</sup>

(FD) If there is reason to feel  $\phi$  then there's reason to do that which  $\phi$  disposes one to do.

Thus, for example, if there is reason to be frightened of  $x$ , then there is reason to avoid  $x$ . If there is reason to be bored by  $x$ , then there's reason not to attend to  $x$ . If there is reason to admire  $x$ , then there is reason to praise, reward  $x$ . Though desires too are affective responses, they differ from other feelings insofar as what they dispose us to do is typically specified by way of their object. If I have reason to desire to go for a walk, I have reason to go for a walk. There is no need to specify what the desire to go for a walk characteristically disposes one to do. This is not so, however, with the other feelings. Take the feeling of admiration: the hermeneutics of this feeling may be such as to specify different characteristic dispositions to act, at different times and for different situations and societies. The other feelings are not as proximate to action as our desires. We shall come back to this point later on.

---

<sup>10</sup> I borrow this formula from Skorupski, 1999a, 38.

Some thinkers take it that in order to abridge the distance between affective reasons and practical reasons we need to postulate that there is reason *to desire to do* that which a feeling reasonably disposes one to do.<sup>11</sup> For example, if at a piano performance I reasonably feel admiration for the performer, there is reason for me to *desire* to express admiration rather than simply reason to *express* admiration. I disagree with this thought. That's because in my view the disposition to desire something is typically accompanied by the disposition to see that thing as ultimately desirable or good for one to have. Now consider the concert example again: claiming that there is a reason *to desire* to applaud in this case would entail that applauding would somehow be good for me. If, on the other hand, we simply used FD we would get the following: if I have reason to feel admiration, I have reason to do that which this feeling characteristically disposes me to do, for example, applaud. Of course, in certain cases, I might receive expressive pleasure from applauding and in that respect applauding would be good for me. However, the fact that I reasonably feel admiration should not be conceptually linked to my well-being.<sup>12</sup>

If what I said above is correct, affective reasons may give us practical reasons that are *not* either implicitly or explicitly grounded in our well-being: that's more evidence that radical welfarism cannot be true. Let us refer to these practical reasons as non-welfarist reasons. Note, however, that so far we haven't said anything about the normative weight of this kind of reasons for action. This will be one of our tasks in the next section.

#### 4. Moderate welfarism: a partial defence

---

<sup>11</sup> Skorupski, for example, believes that FD breaks down into FDD and FDF, where (FDD) states that "If there's reason to desire to  $\alpha$  (or to bring it about that  $p$ ), there's reason to  $\alpha$  (to do that which will bring it about that  $p$ )" and (FDF) states that "If there's reason to feel  $\phi$ , there is reason to desire to do that which  $\phi$  characteristically disposes one to desire to do. Thus, for example if there is reason to feel admiration for  $x$ , there is reason to desire that which the feeling of admiration characteristically disposes one to do, say, express admiration (through FDF); and if there is reason to desire to express admiration for  $x$ , then there is reason to express admiration for  $x$ . See Skorupski, 1999a, 131.

<sup>12</sup> Alternatively, someone might want to say that unless there is no desire to applaud, the applauding can never take place. Assuming that this empirical thesis was true, one might reply by postulating a desire instrumental to triggering the action, rather than an intrinsic desire for the action. We should once more recall the distinction between desires as 'passive sensibility' and the other senses of 'desire'.



I take the two challenges presented above to be fatal to radical welfarism. The question I would like to examine next is whether a more moderate form of welfarism could still be endorsed. This form of welfarism must acknowledge the two objections just discussed. It must deal with (1) the fact that there are values such as aesthetic value that cannot be understood in terms of well-being; and (2) the fact that there are practical reasons that are not derived from well-being. Finally, to make our task simpler, we will assume for the rest of this chapter that, when deliberating, agents need not take into account the way their actions affect other agents and/or beings liable to have an interest. We will limit our discussion to what there is reason and/or most reason for an agent to do considering that his or her actions have no impact on any other agent or being liable to have an interest. This, of course, is a very strong assumption to make and will be duly dropped in Chapter 9. That is why, even if we were to find a defensible form of welfarism in this section, it would only be a *partial* defence of welfarism.

In light of (1) and (2) above and our assumption, I will put forward the following formulation of moderate welfarism:

(EMW) There can never be most reason for an agent to do something that isn't the best outcome in terms of his or her good.

This rather egoistic formula (the 'E' in 'EMW' stands for 'egoistic') would be welfarist because an agent's own well-being would be the ultimate criterion of practical reason for each agent. The fact that this formulation is egoistic should not worry us here. It is a consequence of our assumption that no third parties are involved in the agent's deliberation, and this assumption will soon be dropped. What we should focus on, however, is how this formulation fares with (1) and (2) above.

#### 4.1. The first challenge

To borrow from Crisp's own examples, a moderate welfarist may want to say that Michelangelo's accomplishments are good because we have reason to respond to these goods by having a certain pro-attitude such as admiration. In this case, we would say that they are *aesthetically* good. However, he would add, having a reason to *feel* admiration towards Michelangelo's accomplishment does not seem in and of itself to give anyone a

reason to *pursue* the kind of accomplishment Michelangelo achieved. The moderate welfarist may very well continue to assert that in the realm of action an agent's own well-being is the only good. The same reasoning applies to the 'grass counter' example: we may not need well-being at all in order to *understand*, *appreciate*, or *judge* how much more significant is a life spent trying to find a cure for cancer rather than a life spent counting blades of grass. The moderate welfarist may however insist that the fact that these pursuits are more valuable in certain respects (say, by being more excellent) does not automatically translate into practical value. There may be more reason to *admire* the life of the scientist, but it does not follow from that that everyone would also have more reason to *pursue* that life himself.

It might be replied that someone other than Michelangelo does not have reason to pursue the same kind of accomplishment as Michelangelo precisely because he or she is not as talented and thus incapable of producing such accomplishment. The really interesting question, however, would have to concern someone as talented as Michelangelo or, for the sake of the example, Michelangelo himself. Did *he* have reason to pursue the aesthetic accomplishment he did pursue? If so, what kind of a reason was it? In order to answer this question we need to know more about the specificities of the case. Suppose Michelangelo (with all his talents) has a choice between two types of life. *Life A* is a life of artistic achievements but without much pleasure and ease; *Life B* is a life of ease and simple pleasures but no great artistic achievements. Remember also our assumption that Michelangelo's well-being aside, each of these two choices yields exactly the same outcome from a third-person point of view. Michelangelo, then, can worry only about the repercussions of his choice on his life. How is Michelangelo to choose? It seems to me that there are two dimensions: the welfarist and the non-welfarist or, in this case, the aesthetic one. There are also three interesting sub-scenarios:

(M) Given the set of facts *C* about Michelangelo, Michelangelo rightly evaluates *Life A* as the best life for him, the one that would most realise his well-being.

(M') Given the set of facts *C'* about Michelangelo, Michelangelo rightly evaluates *Life A* as more valuable from the aesthetic point of view and yet considers *Life B* more valuable from the point of view of his well-being. That is, he might see how a life of artistic achievement may contain something more admirable but he

nonetheless might find *Life B* more desirable. Suppose that Michelangelo is in fact right in his evaluation. Given the set of facts *C'* about himself and his situation, *Life B* is the better life for him to live; there is more reason for him to *live* that life.

(M\*) Given the set of facts *C\** about Michelangelo, Michelangelo rightly evaluates that there is reason for him to *live* the life of achievement precisely because it contains more aesthetic value. The idea is something of this kind: *Life A* is more desirable for him because it contains the pursuit of a more (aesthetically) valuable end.

Firstly, note that these three cases are all compatible with moderate welfarism. Secondly, something might be said about the difference between (M) and (M\*). In (M) Michelangelo might be the kind of person who is not moved in any way by the fact that a life containing aesthetic value is better in any sense than a life not containing any. Michelangelo might rather be the kind of person who wants to paint simply because he finds it pleasant. In (M\*), however, Michelangelo desires to live a life of artistic achievement because artistic achievement is aesthetically worthy.<sup>13</sup> That also points at the difference between (M\*) and (M'). In the latter case, Michelangelo is simply not the kind of person who finds aesthetic *value* to be a desirability feature to be pursued in *one's* projects and life.

Finally, we must wonder what case is left open to non-welfarism. Take case (M'), in which Michelangelo has most reason to desire *Life B*. Suppose someone wanted to claim that *Life A* is 'a more valuable life'. What would that mean? Remember that we are operating under the assumption that from a third person or moral point of view *Life A* and *Life B* are identical. Michelangelo can see that *Life A* contains more aesthetic value. But in what sense would that life be 'more valuable'? We should imagine that this claim involves the idea that aesthetic value is somehow a higher kind of value than pleasure and thus any life instantiating it would be a more valuable life. This claim, however, would be false if it also entailed that *Life A* is more choiceworthy and thus that Michelangelo has more reason to choose that life. On what grounds would he have more reason to choose it? He would be

---

<sup>13</sup> Let us stress that in (M\*) Michelangelo does not think that aesthetic achievement is valuable because it is he who does it. Michelangelo can recognise the value of aesthetic achievement without finding it good for him as a personal endeavour as in (M'). In (M\*) however, Michelangelo finds that a life spent attempting to achieve beautiful work of arts would be good *for him*. In this sense aesthetic achievement is especially valuable to him.

made worse off by this life and no one else but him would be affected by his choice. A non-welfarist, however, might defend the following claim:

(PR) Excellence has practical value as such; it should always be pursued irrespective of its contribution to well-being.

This position, of course, needs a substantive argument. Such arguments are often found in perfectionism. Before entering a discussion of perfectionism, however, let me note that so far I have shown moderate welfarism to be compatible with the existence of a good number of non-welfarist values. The unsolved problem posed by perfectionism, however, is the same kind of problem raised by the second challenge against radical welfarism: there are reasons *for action* that are non-welfarist. The discussion of perfectionism, then, leads us straight to tackle the second challenge.

#### 4.2. The second challenge

(EMW) seems plausible as it allows all kinds of *pro tanto* reasons for actions arising from reasonable feelings. Thus if, say, my feeling of admiration for something is reasonable, I may have a *pro tanto* reason to do that which the feeling of admiration characteristically disposes one to do, say, express admiration by applauding. Yet, on (EMW) I cannot have most reason to do that action if it clearly goes against my interest to do so. Thus, all *pro tanto* reasons for actions are, as it were, placed under the filter of an agent's well-being: could it be shown that they were not in the agent's best interest, then there cannot be most reason for the agent to do them.

Take the performance example once more. I find that the performance was admirable and rightly take it that there is reason to applaud. Yet, suppose that, due to a bad wrist tendonitis, the pain I would feel if I clapped my hands would exceed the expressive pleasure I would get from applauding (if indeed I got any). Surely, there is reason for me to applaud. But it seems intuitively correct to say that that reason is never going to be strong enough to override reasons stemming from my well-being. Some might not agree. If they don't, however, and assuming that my choice has no negative effects on any third parties' interests, what arguments can they give to show that even in this case there is most reason

for me to do that which affects my well-being negatively? At this point we can look at perfectionism as providing an answer to this question. More specifically, let us look at Thomas Hurka's version of perfectionism:

This moral theory [i.e., perfectionism] starts from an account of the good human life, or the intrinsically desirable life. And it characterizes this life in a distinctive way. Certain properties, it says, constitute human nature or are definitive of humanity—they make humans humans. The good life, it then says, develops these properties to a high degree or realizes what is central to human nature. Different versions of the theory may disagree about what the relevant properties are and so disagree about the content of the good life. But they share the foundational idea that what is good, ultimately, is the development of human nature.<sup>14</sup>

Let us think back to (M'): Michelangelo rightly evaluates *Life A* as more valuable from the aesthetic point of view and yet considers *Life B* more desirable from the point of view of his well-being. Hurka's version of perfectionism would require Michelangelo to choose *Life A*, irrespective of Michelangelo's own well-being, if it could be shown that that would most develop Michelangelo's human nature. We have here two distinct normative ideals: the development of human nature and the pursuit of one's own well-being. In this example, I am assuming that the first would give Michelangelo most reason to choose *Life A*, while the second would give Michelangelo most reason to choose *Life B*. More precisely, Michelangelo's choice has to be conceived as one between the following two options: developing his human nature at the price of a loss of well-being, or failing to develop his human nature and do what is most in his interest. Note also that Michelangelo cannot in this example be the kind of person for whom developing his human nature is a part of his well-being.<sup>15</sup> What we have to show on behalf of the moderate welfarist is that, though the

---

<sup>14</sup> Hurka, 1993, 3.

<sup>15</sup> Writers such as McNaughton and Rawling might not accept this last point as they fail to notice that there is a conceptual distinction between welfarism and perfectionism. Look for example at the following passage: "[Welfarism] does not state that what determines whether some action or outcome instantiates some particular value is the bearing of that action or outcome on human welfare. ... [Rather i]n determining what constitutes an achievement we just *are*, in part, determining what it is for a life to go well." And "in determining which qualities or character are excellences we are determining what the good life for humans is." McNaughton and Rawling (2000), 157-158; 158 n. 2. McNaughton and Rawling seem to conflate two things: on the one side, what things can make a life a good life, and on the other, what things enhance a person's well-being or make his life go well for him. Contrary to what they seem to think, though determining an excellence may determine what it is for a life to be a good human life, it is possible for a good human life not to be good for the agent whose life it should be. To claim the contrary, would require a substantive argument. At any rate, one cannot simply assume a conceptual identity between a life's

perfectionist ideal may afford agents with reasons for action, there can never be most reason for an agent to do something that is not in his best interest.

We must remember that well-being is essentially connected to the notion of self-sacrifice: if something is a rational self-sacrifice for  $x$  then it will not be the best option in terms of  $x$ 's well-being and *vice versa*. Now, to say that reasons stemming from the perfectionist ideal can never override reasons concerning our own well-being, is to say that it can never be rational to make a self-sacrifice for the sake of personal perfection. Consider (M'): if under the perfectionist ideal Michelangelo chose *Life A*, a life of lesser well-being, his choice would amount to a self-sacrifice. Is this self-sacrifice rational? Remember that Michelangelo rightly evaluates that *Life A* is not the best life in terms of his well-being. He does not have a *desire* to live this life and does not conceive of it as good *for him* to live. Furthermore, we can also suppose that Michelangelo is aware that no one else would be affected by his choice. If he did not choose *Life A*, though some people may still think less of him, his moral responsibility would not be engaged.<sup>16</sup>

We are now faced with the following substantive issue about reasons: can perfectionist considerations ever make it rational to self-sacrifice? I would want to answer negatively. If upon his perfectionist choice of *Life A*, Michelangelo ended up leading a miserable life, I believe it would be appropriate for anyone to ask Michelangelo to provide some justification for his choice. Why renounce a much better life for yourself? It would be natural to point out to Michelangelo that as a result of his choice no one else would fare better while he would fare worse. It would be natural for many of us to assume that Michelangelo in fact *wrongly* assessed the situation: we would think that after all *Life A* must have been what Michelangelo had most reason to *desire*, as in case (M\*). If, on the other hand, Michelangelo chose *Life B*, the life of well-being, his choice would not be found very puzzling. Michelangelo, we would think, chose what was best for him: what could be wrong with that when no one else was affected by his choice? The perfectionist

---

choiceworthiness and its well-being. In the words of Sumner: "there is no logical guarantee that the most developed specimen will also be the best off, or that their undeveloped rivals would not be faring better". Sumner, 1996, 24.

<sup>16</sup> I use the term 'moral' here to mean that Michelangelo would not be blameworthy for his choice. I will say more about the link between morality and blameworthiness in Chapter 9 §3.2. On a different note, one could say that what is important to Michelangelo is that others as well as himself did not "think less of him." But that would not be part of the example at hand. It would be relevant if dignity or self-esteem was to be

ideal simply seems to lack the substantive normative force to ever justify self-sacrifice. To say that, however, is not to deny that moderate welfarism allows (a) that the perfectionist ideal can afford agents with *pro tanto* practical reasons and (b) that these *pro tanto* reasons could become sufficient insofar as they are compatible with the agent's best interest.

Finally, note that Hurka's perfectionism is intended as a full-fledged moral theory. Hurka writes that the good life is that which develops human nature and that the development of human nature is ultimately good. I take these claims to mean that it is ultimately good that each individual has a good life. As we saw, however, the good life may not be good for the agent whose life it is, as it may involve the agent's self-sacrifice. On this view, then, it can at least in principle be the case that what is ultimately good is something that is not good for any human being in particular, and that would indeed be a counterintuitive moral theory.

## 5. Conclusion

Let us draw some conclusions concerning the first part of our defence of moderate welfarism. This view is weaker than radical welfarism in two important respects. Firstly, it allows for the existence of other values outside the realm of action. We need not appeal to well-being in order to understand, appreciate, or judge the value of a work of art or of any other form of achievement. In this respect, I am in agreement with Scanlon when he claims that well-being is not a "master value". Secondly, even within the realm of action, moderate welfarism allows for non-welfarist practical reasons. Next we tested the truth of the claim that there can never be most reason for an agent to do something that is not compatible with his or her well-being, when no third parties are affected by the agent's choice. We did so by examining a case of conflict between reasons stemming from our well-being and non-welfarist perfectionist reasons. We argued that in the case of conflict between these two types of practical reasons, welfarist reasons seem to always carry greater reason-giving force. Our discussion, however, has been conducted under the assumption that no third parties would be affected by one's choice. In Chapter 9, we shall drop this important assumption.

---

subsumed under the idea of human perfection rather than the idea of well-being though I do not know of any

## Chapter 9      Moderate Welfarism, Morality, and Ethics

### 1. Introduction

I concluded Chapter 8 claiming that there can be practical reasons that aren't grounded in an agent's own well-being, though I argued that there can never be most reason for an agent to do something that is not the best outcome in terms of his or her own good. This conclusion was reached under the assumption that the agent's choice would not affect the interests or well-being of any third individual. In this final chapter, I shall defend moderate welfarism while dropping this assumption. As we shall see, there can be several versions of moderate welfarism. I will defend a universalistic or, more precisely, an impartial version of moderate welfarism that will ultimately be stated as follows:

(MW) There can never be most reason for an agent to do something that does not bring about the best outcome in terms of the impartial good.

The impartial good is a welfarist notion, as I will take it to be a positive function of the well-being of all individuals and nothing else. By 'impartial', I mean that reasons stemming from consideration of the impartial good are both universalisable and agent-neutral: we will call these reasons *impartiality reasons*. In what follows we shall examine the meaning of these terms in greater detail (§2). (MW) is a form of welfarism as it takes it that the impartial good is the ultimate standard of practical reason. Just as (EMW), however, (MW) recognises the possibility of non-welfarist reasons. In this chapter, I will focus on one specific kind of non-welfarist reasons, which I shall call *morality reasons* i.e., reasons that stem from feelings such as blame, guilt, shame and disdain (§3). Morality reasons typically single out our moral duties and obligations, and the actions that are morally wrong, right, permissible and so on. They reflect the deontological principles embodied in a community at a time. As we shall see, they can both be agent-relative and agent-neutral. If (MW) is true, however, there cannot be most reason to act in accordance

---

reason to believe that.



with a morality reason if the 'morality-action' is not what brings about the best outcome under the idea of the impartial good.<sup>17</sup>

It may be objected that morality reasons do not enjoy any rational authority of their own, that it can be demonstrated in all cases that they are reducible to or derived from the impartial good. Thus, an even more direct form of welfarism than that expressed by (MW) would be true. This would be the claim of what I shall call a pure form of teleology, a position I intend to resist in two ways: firstly, by claiming that judgements about morality and judgements about the impartial good stem from different epistemologies (§4); secondly, by claiming that if moral reasons were reducible to reasons stemming from the impartial good, where this was understood as nothing other than the positive function of the well-being of all individuals, then it must be the case that moral notions such as fairness are components of individual well-being. I will claim that fairness is not a component of individual good (§5).

These two arguments should incline us to accept the conclusion that morality reasons are independent from the impartial good, that they cannot be derived from it. To say that, however, is not to say that they are not correctible by it (§6). I will defend a conservative holistic position: morality reasons are embodied in a given social-historical context. The practices they involve partly constitute the ethical life of agents in a determinate society. To say that, however, is not to say that they cannot be submitted to rational scrutiny from the standpoint of the impartial good. Finally, I will examine how (MW) fares with the possibility of supererogatory actions, with Bernard Williams's charge of the dislocation of ethical life, and with a charge coming from the rational egoist camp (§7).

## **2. From the good of one to the good of all: the impartial or general good**

In this section, I shall introduce the notion of the general good. We must trace at least some of the details of what is involved in the step that takes one from recognising that each

---

<sup>17</sup> To call reasons stemming from consideration of the general good impartiality reasons is not to say that morality reasons cannot display a form of impartiality. What is clear, however, is that impartiality (intended as agent-neutrality and universalisability) is not a defining feature of all morality reasons. As we shall see, some of our moral constraints are typically agent-relative (for each  $x$ ,  $x$  has reason to care for  $x$ 's offspring).

agent's well-being is valuable to him or her, to recognising that well-being is valuable *tout court*. As we shall see, this discussion cannot be conducted without a brief account of the structure of agent-relative and agent-neutral reasons.

### 2.1. Universalisability, agent-relativity, and agent-neutrality <sup>18</sup>

The largest part of this thesis explored the importance of well-being from a first-personal point of view. We concluded that something cannot be part of an agent's well-being unless there is reason for him or her to desire it. Subsequently, we saw how this affective reason yields a practical reason through the bridge principle (FD). We also saw that not all of our reasons to do something stem from our desires. It would not be true to claim that there is reason to do something if and only if there is reason to desire to do that thing. However, it would be true to say that if there is reason to desire something, i.e., if that thing is good for an individual, then, through (FD), there is reason for that individual to pursue that thing. This would be the thesis that, through (FD), well-being affords *agent-relative* practical reasons:

(WB) ( $y$ ) (If something is for  $y$ 's good, then  $y$  has reason to pursue it)

where  $y$  ranges over persons, i.e., agents with a capacity to think and act on reasons. An important feature of an agent-relative reason such as (WB) is that it is universalisable. Had it not been universalisable (WB) might have looked as follows:

(IE) ( $y$ ) (If something is for *my* good that gives  $y$  reason to pursue it)<sup>19</sup>

Every person  $y$  has reason to desire and pursue *my* well-being. This thought would be non-universalisable because it contains the non-incidental rigid designator '*my*', in the reason-predicate: if it were eliminated the fundamental thought of the sentence would be lost. It is important to notice here that I use the term 'universalisable' in such a way that a universally quantified principle which nonetheless contains a non-incidental rigid

---

<sup>18</sup> The views expressed in this subsection are influenced by Mackie, 1992, Chapter 5 and Skorupski, 1999, Chapter 3.

<sup>19</sup> Sidgwick referred to what we here formulated as principle (IE) as to "irrational egoism". Sidgwick, 1907, 420-421.

designator is *not* 'universalisable'. *Universalisability* should not be confounded with *universality* understood as the following thesis:

(UT)  $(x) (y) (\phi)$  (if under circumstances C,  $x$  has a reason  $R$  to  $\phi$ , then, under circumstances C,  $y$  has a reason  $R$  to  $\phi$ )

Here,  $x$  and  $y$  are both agents and could be identical, and  $\phi$  stands for the set of actions open to both  $x$  and  $y$  under circumstances C. (IE) is not universalisable even though it complies with the universality thesis (UT). Universalisable principles can always be expressed in impersonal terms while that is not necessarily the case for universal principles. Principle (IE), for instance, cannot be recast in impersonal terms. If I uttered (IE) and (IE) were true then everyone would have reason to pursue Rodogno's well-being. That is not the case for (WB): the well-being to be promoted here is not necessarily mine. Each individual's well-being gives him or her reason to act so as to promote it: on (WB) we cannot say that Rodogno's well-being is valuable because it is Rodogno's. We might at most say Rodogno's well-being is valuable to him because it is his, and just as we can say this for his case we can extend it to all other agents.

To say that (WB) is universalisable, however, is not yet to accept that well-being is an agent-neutral value. Universalisability is not what distinguishes agent-neutral from agent-relative reasons. What would an agent-neutral reading of (WB) look like?

(U)  $(x) (y) (\phi)$  (if  $\phi$ -ing enhanced  $x$ 's good that would give  $y$  reason to  $\phi$ )

For every person  $x$ , and every person  $y$ , and every action  $\phi$  open to  $x$  if it were the case that  $\phi$  enhanced  $x$ 's well-being that would give  $y$  reason to  $\phi$ . This is an agent-neutral reason. However, (U) could not possibly be the agent-neutral construing of (WB), as it expresses quite a different thought from (WB). According to (U), the reason-giving fact is that the action in question would enhance the well-being of *someone*. In the case of agent-relative reasons such as (WB), however, the reason giving fact is precisely that the well-being that

stands to be enhanced is the *agent's own*.<sup>20</sup> If the indexical (*his* or *her*) in the reason predicate cannot be eliminated without changing the sense of the sentence, then we are in the presence of an agent-relative reason. Agent-relative reasons cannot simply be construed as agent-neutral reasons.<sup>21</sup> The reverse is also true. In general, any agent-neutral reason-predicate,  $R\phi$ , can be converted into an agent-relative predicate,  $\phi$  is an  $R$ -ing by  $y$ . That, however, does *not* show that it expresses an agent-relative reason.<sup>22</sup> Agent-relative and agent-neutral reasons, then, are mutually exclusive.

## 2.2. The impartiality thesis

This discussion has an important bearing on our thesis. So far we have defended the claim that well-being has agent-relative practical value.<sup>23</sup> It is open to each agent then to wonder why his or her well-being would count more than anyone else's. My well-being, one could think, is as important to me as anyone else's well-being is important to him or her. So, for example, just as I have good reason to desire to do what enhanced my well-being, everyone else would have reason to desire to do what enhanced his or her well-being. If an agent grasped this much, then he would grasp the *universalisability* of agent-relative well-being reasons. The agent who grasps this thought, however, might be able to reason a step further. He might see how there is reason for him not to take any action likely to affect someone else's well-being negatively, precisely because he sees that each agent's well-being is valuable to him or her. Once again, to realise this would amount to realising that this reason is universalisable. However, this time the agent will have embraced more than

---

<sup>20</sup> Note also that (IE) is also agent-neutral but not universalisable. On (IE) everyone has reason to promote my well-being. The reason-giving fact is that it is *my* well-being, not that the well-being that is being promoted is relative to the agent in question.

<sup>21</sup> *Contra* Nagel, 1971, Chapter X. Note that in this work, Nagel refers to agent-relative and agent-neutral reasons as respectively subjective and objective reasons.

<sup>22</sup> The following example will illustrate this point. Compare:

- (N)  $(y, \phi)$  (if  $\phi$ -ing increased the number of needy people being relieved, then  $y$  has a reason to  $\phi$ )  
 (T)  $(y, \phi)$  (if  $\phi$ -ing is the relief of needy people by  $y$ , then  $y$  has a reason to  $\phi$ )

(N) is definitely an agent-neutral reason. In (T) we can find an occurrence of an agent variable in the reason predicate but that does not show that the *reason* expressed by the reason predicate in (T) is agent-relative. For (T) has exactly the same force as (N) in which the reason is agent-neutral. (N) and (T) are analytically equivalent. "There is no possible world in which one of them is true and not the other. And it's a priori that there is not." Skorupski, 1999, 53.

<sup>23</sup> Remember that we endorse the buck-passing account of goodness: this allows us to say in this case, that if there is agent-relative reason to desire  $o$  then  $o$  has agent-relative *value*, or is valuable relative to an agent.

simple universalisability. He would not be thinking that enhancing other people's well-being is good insofar as it is good for him. Rather, what would motivate his actions would be the idea that well-being is valuable irrespective of the person whose well-being it is. This is the idea of impartiality. This idea is expressed by (U), the thesis that well-being provides universalisable agent-neutral reasons.

If each individual's well-being has some weight, then each agent has reason to consider how the well-being of each individual is affected by his or her actions. This last point brings in the idea that (U) involves an impartial positive function of the well-being of all individuals, which we will henceforth interchangeably call the good of all, the impartial or general good, or the good *tout court*. The function is impartial in the sense that it is both agent-neutral and universalisable: it takes everyone's well-being into account and gives no one individual special weight. It should also be noted that this function is neutral with respect to a number of impartial distributive principles. Egalitarians could accept (U) while arguing that there is reason not to enhance an individual's good if that diminishes equality. (U) is also compatible with the principle of Pareto-optimality: if the well-being of an individual is improved and the well-being of no individual is diminished, there is an increase of the general good; to improve that individual's well-being is a good thing. In turn, the Pareto principle is compatible with a number of distributive principles such as aggregate well-being maximisation,<sup>24</sup> leximin<sup>25</sup> distributions, and threshold principles.<sup>26</sup> It is not the task of this thesis to argue in favour of any of these principles in particular. Rather, in this section, I shall only localise the level of investigation at which the question of what principle should be favoured is to be discussed and the methods that should be used to solve this question.

---

<sup>24</sup> Note that aggregate maximisation would have not been compatible with the Pareto Principle, had the latter been formulated so as to express the necessary as well as the sufficient conditions for something to count as an increase in the general good.

<sup>25</sup> For example, Rawls, 1971, 83: "first maximize the well-being of the worst-off representative man; second, for equal welfare of the worst-off representative, maximize the welfare of the second worst-off representative man, and so on until the case which is, for equal welfare of all the preceding n-1 representatives, maximize the welfare of the best-off representative man."

<sup>26</sup> For example, Skorupski, 1999, 90-91: "... the principle which I call *Threshold Justice* ... holds that one should maximize aggregate utility subject to a threshold below which no individual is allowed to fall. ... Unlike leximin, Threshold Justice does not proscribe any improvement, however massive, to someone's position when it is offset by a deterioration, however small, in the well-being of someone less well-off. And unlike the principle of aggregate utility, it does not allow indefinite worsening of a person's position so long as that is offset by compensating gains of well-being to others."

The discussion concerning the shape of the impartial function of the well-being of all individuals, is not directly concerned with theories of justice. Someone involved in our discussion would not be in the process of determining what institutions, actions, practices, and characters would yield the most just society. He would be determining a less practical principle; in the words of Mill, he would be dealing with the ‘ultimate standard of practical reason’. Of course, whatever principle might be agreed upon at this level would provide some of the guidelines informing our decisions at the more practical level. Determining the shape of the ultimate standard of practical reason is a process intended to determine the truth of a type of normative claim. In Chapter 4, we sketched the dialogical epistemology involved in determining the truth of this type of claim. As we mentioned there, there is no guarantee that the reasonable inquirers involved in this process would ever converge on one single principle. Importantly, however, the attempt to define the impartial good cannot be conducted in a vacuum; it must be informed by empirical facts about human beings and about the form of their moral and social interactions. Specialised knowledge of the human and social sciences might be required.<sup>27</sup>

### 3. Morality the institution

Sometimes we might find ourselves in a situation in which we can see that there is most reason to act in a certain way though it is clear that we are under no moral obligation to do so. The best action, as it is sometimes put, might go *beyond* the call of duty. The possibility of supererogatory actions offers initial plausibility to an important distinction, namely, that between what we will call ‘morality the institution’ or ‘the system of morals’, on the one side, and ‘morality conceived as pure practical reason’ or ‘morality conceived from the standpoint of pure ethical theory’, on the other. Below, I will refer to the reasons stemming from each of the elements of this distinction as respectively, morality reasons and impartiality reasons or considerations stemming from the impartial good. I have already said something about the impartial good. In this section, I will explain what kind of reasons we take morality the institution to afford.

---

<sup>27</sup> The importance of this remark will become fully clear in §7.2.

### 3.1. Morality reasons and the feeling of blame

Different philosophers have attempted to characterise morality the institution, i.e., the set of practices embodied in the customs, the laws, and the traditions of a community understood positively or ideally, by reference to the blameworthy. Mill, for one, wrote:

We do not call anything wrong, unless we mean to imply that a person ought to be punished in one way or another for doing it... It is a part of the notion of Duty in every one of its forms, that a person might rightfully be compelled to fulfil it. Duty is a thing that may be *exacted* from a person, as one exacts a debt. Unless we think that it might be exacted from him, we do not call it a duty... There are other things, on the contrary, which we wish that people should do, which we like or admire them for doing, perhaps dislike or despise them for not doing, but yet admit that they are not bound to do; it is not a case of moral obligation; we do not blame them, that is, we do not think that they are proper objects for punishment. ... I think there is no doubt that this distinction lies at the bottom of the notions of right and wrong...<sup>28</sup>

Someone cannot be held to have done something morally wrong, to have failed to do that which she was morally obligated to do, unless she is an appropriate object of blame and punishment. More recently, both Gibbard and Skorupski have taken up and developed Mill's view.<sup>29</sup> One noticeable difference between Mill on the one hand and Gibbard and Skorupski on the other is the way in which being 'to blame' is understood. For the latter, being to blame is understood in terms of what it would be rational for other people to *feel* about the agent. Mill, on the other hand, seems to shift between someone being worthy of the blame feeling and it being rational for someone to blame the agent where that refers to the action of blaming. In Millian fashion, however, both Gibbard and Skorupski would roughly agree with the statement that:

(B)  $\phi$  is morally wrong if and only if the agent ought to be blamed for doing  $\phi$ .<sup>30</sup>

---

<sup>28</sup> Mill, 1998, 93, V14.

<sup>29</sup> Gibbard, 1990, Chapter 1, 40-5; Chapter 7. Skorupski, 1999a, Chapters II and VII.

<sup>30</sup> Gibbard would actually say: "An act is *wrong* if and only if it violates standards for ruling out actions, such that if an agent in a normal framework of mind violated those standards because he was not substantially

As just explained, here the blaming refers not to the action but to the feeling. The other side of the blame feeling is guilt. Thus, the agent who should reasonably be blamed for an action should feel guilt. The system of morals, then, commands us to follow a set of rules or principles that can be expressed as both agent-relative and agent-neutral reasons for action stemming from the blame feeling. Principle (K) should do as an example of an agent-neutral moral reason:

(K)  $(y) (\phi)$  (if  $\phi$  is the killing of an innocent person, that gives  $y$  reason not to  $\phi$ ).<sup>31</sup>

that is, there is reason for everyone not to kill an innocent person. As for a typical example of an agent-relative moral reason we may look at (F):

(F)  $(y) (\phi)$  (if  $\phi$  is the caring for  $y$ 's child, that gives  $y$  reason to  $\phi$ ).<sup>32</sup>

that is, there is reason for everyone to care for his or her children. To say that (F) and (K) are moral reasons is to say that were it the case that respectively  $\phi$ -ing and not  $\phi$ -ing were open to  $y$  and  $y$  failed to act accordingly then  $y$  would be blameworthy and has reason to feel guilty. It is important to explain here what we mean by saying that  $\phi$  is an act *open* to  $y$ . By that it is implied that on this occasion  $\phi$ -ing was within  $y$ 's physical power, or that  $y$  can be held responsible for failing to *realise* that  $\phi$ -ing would yield  $y$ 's killing of an innocent person or  $y$ 's failure to care for his or her child. In certain circumstances, a cognitive deficiency such as inattention cannot be appealed to to show the non-openness of  $\phi$ -ing to the agent. When, however, an action is *not* open to an agent in this way then it would be unreasonable to blame someone for doing or failing to do that action. If someone surreptitiously hypnotised me, put a loaded gun in my hand and ordered me to shoot an

---

motivated to conform to them, he would be to blame. To say that he would *be to blame* is to say that it would rational for him to feel guilty and for others to resent him." Gibbard, 1990, 45. As for Skorupski, he would stress that (B) is not intended to be a *strict definition* of moral wrongness. In fact, he takes wrongness to be the semantically primitive term, though liable to some forms of conceptual analysis. He would describe (B) as a *characterisation* of the concept of moral wrongness. See Skorupski, 1999a, Chapters II and VII.

<sup>31</sup> For each agent  $y$  and each action  $\phi$  open to  $y$ , if it were the case that  $y$  is the killing of an innocent person, that would give  $y$  reason not to  $\phi$ .

<sup>32</sup> For each agent  $y$  and each act  $\phi$  open to  $y$ , if it were the case that  $\phi$  is the caring for  $y$ 's child, that would give  $y$  reason to  $\phi$ .



innocent passer-by, it would not be reasonable for someone else to blame me and for me to feel guilty, though of course it would be reasonable for me to feel regret.

The moral practices of a society can be revised from within. The process is holistic: suppose that at time  $t$  it is *not* reasonable to blame the perpetrators of the  $\phi$ -ing practice. However, progressively, reflection and new information reveal that the spirit of that practice is in open contradiction with a large number of other moral practices, the breaching of which is reasonably blameable. No external substantive criterion needs to be appealed to in order to revise the  $\phi$ -ing practice. The internal coherence of the morality system as regulated by the blame feeling would suffice. Or again, suppose that under a certain general state of knowledge within a society it is considered reasonable to feel the blame feeling for a certain type of action. Some new psychological findings, however, show that agents performing that action are typically incapable of refraining from it. They lack the psychological power to do so. Our moral attitudes towards this practice should reasonably be altered.

### 3.2. Morality reasons and the feeling of shame

Some philosophers, however, think that the blame feeling cannot alone characterise the morality system in the way just described. Rawls, for example, writes that:

All by themselves, a morality of shame or of guilt is but a part of a moral view. ... Guilt and shame, remorse and regret, indignation and resentment, either appeal to principles belonging to different parts of morality or invoke them from contrasting points of view. An ethical theory must explain and find a place for these distinctions, although presumably each theory will try to do so in its own way.<sup>33</sup>

Williams, offers the following analysis of guilt and shame:

What arouses guilt in an agent is an act or omission of a sort that typically elicits from other people anger, resentment, or indignation. What an agent may offer in order to turn this away is reparation; he may also fear punishment or may inflict it on himself. What

arouses shame, on the other hand, is something that typically elicits from others contempt or derision or avoidance. This may be equally an act or omission, but it need not be: it may be some failing or defect. It will lower the agent's self-respect and diminish him in his own eyes. His reaction... is a wish to hide or disappear, and this is one thing that links shame as, minimally, embarrassment with shame as social or personal reduction. More positively, shame may be expressed in attempts to reconstruct or improve oneself.<sup>34</sup>

It is not my task here to examine the appropriateness of Williams's analysis of these two moral feelings. It is important, however, to mention something else Williams has to say about the importance of guilt and shame relative to certain societies, namely that of Homeric Greece. Williams notes that in that particular society the contrast between these two feelings wasn't in fact so sharp. Shame seemed to cover all or almost all the ground that guilt covers for us today. Williams then considers and rejects the claim that the Greeks' concept of shame was the same concept as our concept of guilt:

Even though some reactions in those societies were structured in the same way as our reactions of guilt, they were not simply guilt if they were not separately recognised as such; just as shame is not the same when it does not have guilt as a contrast. What people's ethical emotions are depends significantly on what they take them to be. The truth about Greeks societies, and particular the Homeric, is not that they failed to recognise any of the reactions that we associate with guilt, but that they did not make of those reactions the special thing that they became when they are separately recognised as guilt. ... One thing that a marked contrast between shame and guilt may express is the idea that it is important to distinguish between "moral" and "nonmoral" qualities. Shame itself is neutral on that distinction...<sup>35</sup>

Williams goes on to suggest that in our present day society moralists have artificially put too much emphasis on the distinction between the moral and the nonmoral.<sup>36</sup> As interesting as these issues may be, it is not my task here to discuss them in further detail. I need not determine here whether guilt and blame cannot, when taken alone, provide a full

---

<sup>33</sup> Rawls, 1971, 485.

<sup>34</sup> Williams, 1993, 89-90.

<sup>35</sup> Williams, 1993, 91-92.

<sup>36</sup> "It is said that we make a lot of the distinction between the moral and the nonmoral and emphasise the importance of the moral. But how far, and in what ways, is this really true of our life, as opposed to what moralists say about our life? Do we even understand what the distinction is, and how deep it really goes?" Williams, 1993, 92.

characterisation of morality, as suggested by Rawls. Nor will I attempt to determine how much truth there is in Williams's suggestion that moralists overestimate the importance in our ethical life of guilt and its emphasis on the distinction between the moral and the nonmoral. This discussion, however, is relevant to our purposes in the following way. Even if morality the institution could not fully be characterised in terms of the blame feeling and feelings such as shame were relevant to its fullest characterisation, my point would still hold: *the source of morality the institution rests within the affections*. That has an important consequence. Our morality reasons belong in the first instance to affective rather than practical reason. In a manner similar to the reasons grounded in our own well-being, the *practical* import of morality reasons is mediated by bridge principle (FD). We shall soon see the importance of this claim.

#### 4. Teleology, deontology and the epistemology of their reasons

Taking stock of the nature of our morality reasons as illustrated in the last section, I can now begin to show the impossibility of *deriving* our morality reasons from the impartial good. This task will occupy this as well as the next section. This terminology (e.g., "deriving our morality reasons from the impartial good") refers to a fundamental debate in ethics that needs a short preamble.

##### 4.1. Pure deontology and pure teleology

The discussion concerning the relevance or importance of well-being in ethical theory would be very much simplified if it could be shown that welfarism was true because all our reasons, including our morality reasons, are in fact simply derived or deduced from the general good. Such a view we will call pure teleology, where by teleology I mean the theory of ends, the theory of what is desirable either agent-relatively or agent-neutrally. Pure teleology claims precisely that the morally obligatory is *derived* from (or stands in some relation of dependence to) the good. At the opposite pole from this view we find what we may call pure deontology, where by deontology I mean the theory of duty, the theory of the morally right and wrong action, of the morally obligatory, the morally

forbidden and permitted; deontological principles are principles of *morality*. Pure deontology claims that the morally obligatory is prior to the good and the bad and is most clearly stated by Kant:

the concept of good and evil must not be determined before the moral law (for which, as it would seem, this concept would have to be made the basis) but only (as was done here) after it and by means of it.<sup>37</sup>

The position I will henceforth defend does not correspond to either one of these views. I believe it is mistaken to conceive of the relation between these two conceptual categories as one of dependence of the one on the other. In fact, both the morally obligatory and the good are independent from one another, though as I will attempt to show judgements about the good can be appealed to in order to criticise, inform, and correct judgements about the morally obligatory.

In this section, I will attempt to show that morality reasons do not share a common epistemology with impartiality reasons. More precisely, I will claim that judgements about morality reasons stem from the feelings while judgements about the impartial good don't. The distinction in the epistemology of these claims points at the idea that these types of claims are about different types of things, with different meanings. It is not immediately clear to the mind that judging that a state of affairs brings about the most good *is* judging that that is what we are morally obligated to bring forward. It is on the other hand much more intuitively clear how a justified feeling of guilt identifies someone's failure to act on a moral obligation.

#### 4.2. Different epistemologies

I presented morality the institution as the set of practices embodied in the customs, the laws, and the traditions of a society. I suggested that the feelings of blame and guilt, and/or of shame and disdain, were at the basis of our morality judgements. Recall now our discussion concerning spontaneity in Chapter 4. There, I claimed that to acknowledge that

there is reason for a person to feel something is to reflectively accept a spontaneous impulse of one's affections. In the case of well-being, it is to reflectively accept one's desires as rational. Similarly, for our moral judgements: to say that an action is wrong is to reflectively endorse one of the moral feelings just mentioned. These moral affections are the primitive criteria in establishing judgements about morality. That also explains why a bridge principle (FD) is required in order to mediate from the affections to action.

Contrary to morality reasons, the affections are not the primitive criteria for judgements about impartiality reasons. Rather, using a somewhat Kantian terminology rid of its transcendental underpinnings, I would say that our pure willing is such a criterion. Our 'pure' willing is here understood as what one is disposed to will when one does away with all emotions or affections. I would claim that what one is disposed to will unconditionally in this way is an impartial function of the good of all. To acknowledge that there is reason to act in a certain way in certain circumstances irrespective of what one takes oneself to have reason to desire or to feel, is to reflectively accept a spontaneous impulse of one's will: the content of this impulse is embodied by the general good. Judgements about what one has reason to will, then, do not involve the feelings and hence do not require a bridge principle such as (FD).

Some may take it to be farfetched to think that acting in accordance with the impartial good is what one would spontaneously will when no affections are involved. My reply consists in a rehearsal of the thought process that takes rational agents from recognition of the agent-relative value of well-being to recognition of its agent-neutral value. The starting point is the realisation that we all have reason to pursue our own well-being: I can see that my well-being is valuable to me just as yours is valuable to you. From there, an agent may move further and recognise that each agent's well-being is impartially valuable: my well-being is just as valuable as everyone else's, therefore there cannot be most reason for me to do an action that doesn't bring about the best outcome when considering the well-being of all. A state of affairs involving greater impartial good is better than one involving less; there is then more reason to pursue it even if that involved a loss in my well-being. This realisation does not involve the affections but proceeds purely from rationality.

---

<sup>37</sup> Kant, 1996, *Critique of Practical Reason*, 190, (5:63).

To say that, however, is not to say that this process involves no heteronomy. The feelings do play a role in this picture. It must be remembered that the general good is a positive function of each individual's well-being. But well-being and the practical reasons it yields involve our affective side. Someone with no capacity to feel, and more specifically, someone with no capacity to desire, would not be able to find any instance of well-being intelligible, let alone a positive function of the well-being of all individuals.<sup>38</sup> In other words, one cannot begin to appreciate the agent-neutral value of the general good, unless one has the capacity to feel and desire and thus understand the agent-relative value of each person's well-being.<sup>39</sup> Importantly, the same kind of point can be extended to morality reasons. If morality and its reasons stem from the feelings, someone with no capacity to feel cannot find this type of reasons intelligible; in fact he or she would not find any particular instance of (FD) intelligible.<sup>40</sup>

Finally, let me say a word about motivation. Recall the discussion concerning the normativity of well-being in Chapter 3. There, I claimed that for any agent to have the capacity to be motivated by the description of a state as one that is good for her, she must also have the linguistic/intellectual capacity to understand what it is for something to be good for her. To say that an agent has both capacities, however, is not to say that she will unfailingly be motivated by each and every judgement about her well-being, as that would exclude the possibility of acting against one's better judgement. The same applies to the general good. An agent may recognise that something is good *tout court* and yet fail to be motivated to act accordingly.

In conclusion, morality and its reasons are separate from the general good and its reasons insofar as their epistemic sources are different. The primitive criteria for our judgements about morality are our moral feelings, while the primitive criterion for our judgements about the impartial good is our pure willing. The distinction in the epistemology of these

---

<sup>38</sup> To realise that the impartial good involves heteronomy in this way is crucial in avoiding a charge that has often been put to Kant, namely, that *pure* practical reason is an 'empty formalism'. The charge applies to Kant if it is understood as meaning that the attempt to derive moral obligation from the a priori idea of a rational requirement (the universality of reasons) as such is empty. For a discussion of this point, see Wood, 1990, Chapter 9.

<sup>39</sup> It would be interesting to compare this claim with G. E. Moore's idea that something must be good *absolutely* in order for it to be good *for* someone. See Moore, 1903, 99.

<sup>40</sup> Another facet of the 'empty formalism' charge against Kant is thus uncovered: the categorical imperative cannot yield a substantive doctrine of moral duties without a given context of 'ethical life'. These two sides

claims indicates that these types of claims are about different types of things, with different meanings. It is not immediately clear to the mind that identifying the *best* state of affairs *is* identifying what we are morally obligated to bring about. It is on the other hand much more intuitive to think that a justified feeling of guilt identifies someone's failure to act on a moral obligation. The fact that we directly rely on our feelings to know what our moral obligations are while we do not do so in order to find out about the impartial good should explain this divide. If this is true, then, the pure teleologist cannot claim that morality is simply derived from the good. Importantly, however, our position does not entail that the impartial good has no bearing whatsoever in determining the meaning of "wrong". This fundamental point will be developed only after having discarded another attempt to show that morality can be derived from the good.

### 5. What things the impartial good cannot consist of

A pure teleologist might be unimpressed by our argument from epistemology. He might argue that the notions constituting the morality systems, notions such as right, wrong, rights, duty obligation, fairness, and desert, are rather unclear notions and that is precisely why they need to be *grounded* on a clearer *foundation* such as the impartial good.<sup>41</sup> In order to make this claim, however, the pure teleologist must be able to defend the claim that all of these concepts can be captured by the notion of the impartial good. John Broome and Amartya Sen both seem to defend precisely that claim (though as we shall see they might refuse to take on board the foundationalist rationale I presented as backing this kind of objection). In what follows, I will focus on one moral notion in particular, namely, fairness, and see whether it can indeed be reduced to or included in the notion of the impartial good. Success in showing that it cannot reinforces our thesis concerning the impossibility of *deriving* moral notions from the good.

---

of the empty formalism charge give reason not to consider pure deontological ethics at least in their Kantian form as a valid alternative.

<sup>41</sup> I do have someone like Sidgwick in mind. See Sidgwick, 1907, Books iii and iv.

### 5.1. Broome's fairness

Broome considers the following example taken from Diamond.<sup>42</sup> There is a kidney available for transplant, and two people,  $P$  and  $Q$ , who need it to survive. The choice facing a hospital administrator is between tossing a coin to decide who is to get the kidney (alternative  $A$ ) and giving it directly to  $Q$  (alternative  $B$ ). It may be plausible to suppose with Diamond that choosing randomly, e.g., by tossing a coin, would be best as it would give each person a 'fair shake'. It is important to mention that Broome takes the example not to involve any feelings on behalf of the patients. For example in  $A$  we cannot postulate that the belief that  $Q$  is not being treated fairly would cause  $Q$  to have the negative feeling of disappointment. Each person will either receive a kidney or not, without explanation. The process of choice cannot affect their feelings. If the preference for  $A$  is rational, then according to Broome,  $A$  is better than  $B$ . What makes it better is precisely the fact that  $A$  possesses the property of being a fairer outcome. At this point Broome qualifies his position as follows:

My term 'outcome' may be a little misleading. It does not refer particularly to the results of an act rather than the act itself. The outcome in a state of nature includes everything that happens in that state, both before the state is revealed and after. It includes history. If alternative  $A$  is chosen rather than  $B$ , and state heads comes up, then the outcome includes the fact that  $A$  is chosen. If  $A$  is fair, its fairness will be a part of the value of the outcome.<sup>43</sup>

Taken in isolation there is nothing wrong with Broome's claim that the value of an outcome may include the unfairness of a choice. This claim, however, becomes problematic when is taken in conjunction with another claim, one that both Broome and I endorse. The impartial good, we saw, is a positive function of the well-being of each individual and nothing else. Broome endorses this same thesis, which he calls the *Principle of Personal Good* and which he states as follows:

---

<sup>42</sup> The following discussion is in Broome, 1991, 111-115.

<sup>43</sup> Broome, 1991, 114.



(a) Two alternatives are equally good if they are equally good for each person. And (b) if an alternative is at least as good as another for everyone and definitely better for someone, it is better.<sup>44</sup>

Given this principle, if alternative *A* is better, then it must be better for *P*; in other words *P*'s well-being or personal good would be enhanced if *A* obtained. We have been told, however, (1) that what makes the alternative better is precisely the fact that it is fairer, and (2) that no feelings of any kind were involved in the example. It follows then that fairness or being treated fairly is what increases *P*'s well-being; it is part of his well-being or personal good.<sup>45</sup> What we should ask now is whether fairness can in fact be a part of a person's well-being.

Being treated unfairly typically carries with it diminution of other aspects of a person's well-being. When we are treated unfairly, our well-being is normally decreased to the extent to which being treated unfairly deprives us of some other good which we might have otherwise attained. Taking the kidney example, in the unfair alternative *B* it is clear that *P* is deprived of his well-being in all its components, as he is deprived of all his chances to survive. In alternative *A*, on the other hand, he has a fifty percent chance to maintain his well-being with all its components. To take another example, a panel unfairly choosing a less qualified candidate for a job might deprive the other candidate of his chance to pursue his projects and achievements; gaining knowledge of the unfair treatment may cause disgust and great disappointment in the qualified candidate and, to that extent a loss of well-being; but keeping the truth hidden from him may also amount to a loss of well-being insofar as knowledge about one's situation is part of a person's well-being. The fact that being treated unfairly typically carries with it diminution of other aspects of a person's well-being is indicative—though not a sufficient argument—that fairness is not itself a part of well-being.

Let us then try to isolate the putative good of 'being treated unfairly' from the other categorial ends through a thought experiment. Imagine two possible worlds *W* and *W'*: in each of the worlds Rea enjoys precisely the same amount of each of the other goods

---

<sup>44</sup> Broome, 1991, 165.

<sup>45</sup> Elsewhere Broome explicitly writes: "Unfairness, as I have described it, is plainly an individual harm." Broome, 1991, 198.

constituting her well-being, but in  $W$  she is treated more fairly than in  $W'$ . It does not seem to me that Rea has prudential reason to choose to live in  $W$  rather than  $W'$ . That should be clearer when the following is considered. Suppose that *ceteris paribus*, her well-being in  $W'$  is slightly increased. It would seem now that she has good prudential reason to choose to live in  $W'$  rather than  $W$ , though that is the unfair world.<sup>46</sup>

Judgements about fairness cannot be represented as judgements about what an agent has reason to desire for itself though, as we just saw, being treated unfairly often produces loss in well-being. Rather, judgements about fairness are moral judgements, judgements about what is morally obligatory. Like many other morality judgements, they often enter our deliberation as constraints and not as what would maximise utility or the value of outcomes.<sup>47</sup> Suppose that the best outcome is the outcome including the fewest nonviolations of rights because that would be the fairest outcome. Suppose that Outcome  $X$  is the best outcome but that bringing about  $X$  involves the violation of someone's right and there is a moral constraint  $C$  not to do so, as violating someone's right would be unfair to this person. If this moral constraint is to be taken seriously at all then one cannot bring about  $X$  even though  $X$  had been postulated to be the fairest outcome.

This example should show that the fairness of an action, a process, a choice or a state of affairs is not the type of property that lends itself to the teleological rule that more is better than less. What one could at most say is that more is better than less subject to a constraint, namely, that no one is treated unfairly. To admit this much, however, is to admit that fairness cannot be derived from the goodness of outcomes; if fairness is what makes an outcome good, then it would be true that more fairness is better than less. But as we just saw above fairness cannot be understood in this way. This point is reinforced by our previous point that fairness is not a component of an individual's well-being: if we are right in thinking that it isn't, and the impartial goodness of outcomes is a positive function of each individual's well-being and nothing else, then fairness cannot appear as a constituent of the good impartially conceived. Fairness is a concept that belongs to morality and its reasons; it is independent from the good and cannot be derived from it.

---

<sup>46</sup> Similar points are made by Crisp & Moore, 1996, 612; and Skorupski, 1999a, 105-06.

<sup>47</sup> In what follows, I will be rearranging an example taken from Nozick's discussion of side-constraints. See Nozick, 1974, 28-33.

Once again, however, this claim does not exclude the claim that the impartially conceived good has no bearing on our moral reasons including the ones concerned with fairness.

## 5.2. Sen's consequential evaluation

Sen does not take himself to be a foundationalist teleologist of the kind we described at the start of this section. He writes:

It is neither that 'the good' comes first, and then 'rights and duties', nor that rights and duties congeal first followed by the good, but that they are linked concepts that demand simultaneous consideration. While considerations of freedoms, rights and duties are not the only ones that matter (for example, well-being does too), they are nevertheless *part* of the contentions that we have reason to take into account in deciding on what would be best or acceptable to do. The issue surely is *simultaneity*. This differs... from 'sequencing' (beginning, first, with 'the good ... free of deontic notions' and then proceeding to rights and duties)...<sup>48</sup>

Though Sen clearly avoids taking the foundationalist stance we described above, he still believes that the best interpretation of familiar moral notions themselves involves, at least in part, their inclusion as factors that contribute to the value of states of affairs. This view is best seen at work in Sen's treatment of the notion of rights: the value of states of affairs takes into account the intrinsic value of protecting rights and the disvalue of having them violated. Before tackling this claim we shall notice three important features of Sen's 'Consequential Evaluation' of states of affairs. Firstly, the evaluation is intended to be 'comprehensive' insofar as it includes consideration of (a) the final/culmination outcome, (b) the act in itself, (c) the motivation behind the act, and (d) the process by which the motivation was arrived at. Secondly, the evaluation is 'situated': it is impartial but not impersonal, i.e., conducted from the point of view of the agent and not from the point of view of the world. Finally, consequential evaluation avails itself of a maximizing conception of rationality.<sup>49</sup>

---

<sup>48</sup> Sen, 2001, 61. Note that Sen may be taken to use the notions of "what is best" and "what is acceptable" as interchangeable. I can think of cases where what is acceptable is not what is best and *vice versa*.

I have already discussed the impossibility of placing rights understood as side-constraints under the idea of a maximizing rationality, and will not discuss this point further.<sup>50</sup> Another relevant issue arising from Sen's discussion, however, concerns the nature of situated valuations. Sen believes that a good reason for including actions in general and right-violating actions in particular as components of states of affairs in a consequentialist framework is that a system of moral thinking that has this form has the morally appealing property of requiring agents to take responsibility for their actions and their consequences in the proper way. According to Sen, states of affairs have different values when evaluated from different 'positions'. Taking Williams's well-known example of Jim and the Indians, Jim would attach special negative value to *his* killing one of the Indians.<sup>51</sup> What we need to understand, however, is precisely what kind of value is negative in relation to Jim. We should all be in agreement with at least this much, that one's reason not to kill an innocent person is the *agent-neutral* reason (K):

(K) ( $y$ ) ( $\phi$ ) (if  $\phi$  is the killing of an innocent person, that gives  $y$  reason not to  $\phi$ ).

If there are any moral principles at all, then surely (K) is one of them. At least part of Jim's reluctance in killing the Indian is generated by his awareness of being in breach of such a fundamental moral principle. The reason-giving fact as described by (K) is not the killing of an innocent person *by Jim*, but the killing of an innocent person *tout court*. Had the sweat-stained-khaki-shirted captain picked Jim's travel companion, Jack, rather than Jim, then Jack would have been in exactly the same situation as Jim insofar as he would have

---

<sup>49</sup> See Sen, 2000.

<sup>50</sup> See §5.1. in this chapter. It might be objected that an instrumental theory of rights may recognise that rights do not involve absolute prohibitions. According to Scanlon, on this theory, "to claim that people have a certain right is to claim that they have a certain important interest that can be protected only if others' freedom to act is constrained in certain ways. ... But we also recognize that absolute prohibitions against interfering in these ways would be unacceptably costly to others. So, in order to be defensible, rights have to be understood as admitting of exceptions. In deciding whether a defensible right would have to incorporate a certain exception, we must take into account not only the relative seriousness of the interests in question (the interest that the right is meant to protect and the interest that would be sacrificed without the exception) but also the degree to which the secure protection that the right is meant to provide would be undermined by an exception of the kind in question." Scanlon, 2001, 44-5. I would like to make one comment about the instrumental theory Scanlon is putting forward. Suppose that we all have a right to be treated fairly. Scanlon writes that "to have a right is to claim that people have a certain important interest". What would be the particular 'interest' being protected by the right to be treated fairly? Is it some aspect of each individual's well-being? But in that case, the right in question should be directly concerned with the protection of that aspect rather than with fairness. Alternatively, then, 'being treated fairly' may itself be part of each agent's well-being. But that, as we saw above, is implausible.

<sup>51</sup> Smart and Williams, 1973, 98-9. Jim, however, will have to take notice that it is *he* who is also, though less directly, involved in the death of twenty Indians.

been asked to do something that everyone has reason not to do. To realise this much is to realise that the reason not to kill an innocent person is an agent-neutral one. As we saw above, that agent-relative indexicals (the killing *by Jim* of...) can appear in the formulation of this principle is not enough to show that the principle is itself agent-relative. Yet, we should also all agree with Sen (and Thomas Nagel and Williams among others) that there could be a special agent-relative or position-relative disvalue for Jim (and/or Jack) in the killing of this innocent person. What kind of disvalue?

I would claim that this disvalue can only be understood in terms of the agent's self-interest or well-being. Given certain facts about Jim, there could be many ways in which the killing of an innocent person may have special prudential disvalue for him. Jim for example may find the thought of having killed a person extremely disturbing and that unpleasant thought may haunt him for the rest of his life. The whole situation may also strike Jim as one in which the value of his autonomy (another part of his well-being) is being clearly diminished. A choice is forced upon him against his will and the two options open to him are options he would have not himself rationally chosen. Also, Jim may, to a smaller or greater extent, have reason to desire to act virtuously and it may seem to him that the choice he has been offered leaves him no room to act virtuously. Finally, if we changed the example a little, we may imagine that Jim in fact knows and is a good friend of the Indian he is supposed to sacrifice in order to save the others. The thought that it is he who has to take his friend's life may be particularly horrible precisely for the special relation that binds him to his Indian friend.

Sen, of course, cannot agree with our claim that the special disvalue to the agent is to be understood in terms of the agent's self-interest, as that would go against the spirit of the type of moral reasoning that Sen intended his position-relative evaluation to capture. It would be mistaken to explain the wrongness of the killing by appealing to the agent's self-interested objection to the killing. The reason giving fact is that an innocent person would be killed. In this type of case, our moral thinking would tend to focus on the claim of the victim to have his life preserved rather than on any of the putative killer's self-interested objections.<sup>52</sup> Sen might want to reply to our objection by drawing a distinction between the idea of a state of affairs having special disvalue when *evaluated from* a particular position

---

<sup>52</sup> I am indebted to Scanlon's discussion of this point and the one following it. Scanlon, 2001, 47- 8.

and the idea that it is *bad for* the person who is occupying that position. Darwall, for example, claims that there is a:

... difference between what is good from the agent's point of view and what is for her good or welfare.<sup>53</sup>

An agent can of course think that it is good *that*  $x$  obtains without thinking that the obtaining of  $x$  is good for her in any way. I have endorsed this claim throughout this thesis. The problem for someone like Darwall or Sen, however, is that when an agent claims that it is good that  $x$  obtains she is not deliberating from an agent-relative point of view; she *must* think that it is good taking everyone's interests into account. And how can she think that and yet be deliberating 'from the agent's point of view'? Someone might insist, however: there are things that are "good from the agent's point of view" but are not "good for her" and are not good *tout court*. I am not able to think of any one thing fulfilling this description and therefore think that the concept Darwall or Sen are trying to single out is an empty concept. Jim, for example, might think that it is good from his point of view that he does not kill the Indian though that is not good for everyone. However, we put forward a case for thinking that "good from his point of view" is to be read as "good for him". If it were not, then, the disvalue of killing from Jim's point of view would consist in an agent-relative non-self-interested reason not to bring the killing about. This reason, however, would look very much like the moral principle (K) and, as argued above, though (K) may appear to be agent-relative it in fact is an agent-neutral reason.

In conclusion, even if the evaluation of the good and the various moral notions was simultaneous and claimed by Sen to be non-derivative, there are problems about fitting moral constraints into the maximising structure of Sen's consequence sensitive evaluation. What is more, we saw that one of the steps intended to bring in moral notions within consequence sensitive evaluation, namely, situated evaluation, displays features that go precisely against the spirit of customary moral thinking. In the light of the conclusions reached in this as well as in the preceding section, we can look at a non-derivational (and non-simultaneous) way of fitting considerations stemming from the impartial good with our morality reasons.

---

<sup>53</sup> Darwall, 2002b,14.

## 6. The corrective view, conservative holism, and moderate welfarism

### 6.1. Moderate welfarism and the corrective view

The view I would like to defend claims that though our morality reasons and the obligations they place us under cannot simply be derived from consideration of the impartial good, this type of consideration can be appealed to in order to *correct*, *inform*, or *criticise* our moral judgements. Given that morality reasons stem from the feelings, they ought to be mediated by the bridge principle (FD) when entering the realm of action. On the other hand, reasons stemming from the impartial good are principles of *pure* practical reason, unmediated by the feelings: they straightforwardly belong to the realm of action. It is at precisely the stage at which our morality reasons are mediated by (FD) to enter practical deliberation that they meet with considerations concerning the impartial good. It is at that stage that the ultimate standard of practical reason is given the chance to correct the actions stemming from our reasonable moral feelings.

In Chapter 8, we gave the following provisional statement of moderate welfarism under the assumption that no third party would be affected by the agent's choice:

(EMW) There can never be most reason for an agent to do something that doesn't bring about the best outcome in terms of his or her good.

A complete statement of moderate welfarism, i.e., one including morality and impartiality reasons, however, should look as follows:

(MW) There can never be most reason for an agent to do something that doesn't bring about the best outcome in terms of the impartial good.

Let us begin to comment on this principle by saying something about its place in our moral psychology. On our statement of moderate welfarism, it is not the case that for each action an agent has reason to perform, her action needs to be done *out* of the belief that it shall

bring about the best outcome for all. If, however, on one occasion, the agent sees that her action is in breach of the good impartially conceived, she should refrain from doing it. The impartial good works as a constraint.

Next, let us look at the way in which (MW) regulates the interaction between impartiality reasons and other kinds of reasons. Consider reasons stemming from an agent's own well-being first. When the normative impulse accompanying our desire for an object passes the scrutiny of reason, we may correctly take ourselves to have reason to desire that object and, through (FD), we may take ourselves to have *pro tanto* reason to pursue it. If, however, we recognise that reasons concerning the general good tell against the pursuit of that object, there cannot be most reason to pursue it. This does not mean that the initial prudential reason is stricken out of existence; it does not become null. We can still recognise it as a prudential reason and can feel its motivational pull as a reason (not as a mere desire). However, the ultimate standard of practical reason would tell us that it should stand *defeated* by weightier considerations, i.e., considerations for the good of all.

The same kind of picture applies to reasons arising from morality the institution. We saw how it would be plausible to think that many if not all of these reasons can be characterised by the blame feeling. These reasons too stem from affective reason. Just like desires, these feelings lend themselves to a first round of rational validation internal to their hermeneutics. I may feel guilty or ashamed but is it reasonable to have such feeling in these circumstances? Is it reasonable for someone to blame me for something I am in no way responsible for? If on one occasion the answer to these questions were negative, then we would not be in the presence of any moral reason. If the answer were positive, however, then there would be such reason. In that case, through (FD) there may be reason to *do* that which the feeling in question characteristically disposes one to do. If, for example, there is reason to blame someone, then there is reason to express blame, punish him, withdraw recognition, etc., once again in accordance to the norms internal to the blame feeling. Yet, once we enter the realm of practical reason, the reasons to *do* that which the blame feeling characteristically disposes one to do are checked by the constraint of impartiality. If it could be clearly shown that to express blame in that instance was to bring about widespread havoc and misery, for example, then there cannot be most reason to do it.



## 6.2. The corrective view and conservative holism

The corrective view is a form of conservative holism in practical reason. The conservative holist in theoretical reason believes that something like a Cartesian method of enquiry—one that begins by dismissing all of one's previously held beliefs and proceeds by reconstructing a system of beliefs by applying purely rational criteria to pure data—is a non-starter. Similarly, the conservative holist in practical reason holds that the point of departure for reforming a set of practices is always situated within the specific social and historical context in which those practices exist. One should not start by withdrawing from these practices. Given a set of practices and an ultimate criterion of practical reason it is possible to identify, through the criterion, the places at which these practices need revision. Conservative holists believe that systems of practices cannot be revised *en bloc*; their revision can only be piecemeal.

One objection to *en bloc* revision would be a practical one. How could we suspend acceptance of the moral practices of a society *en bloc*? There is, however, also an objection in principle to such revisions. Conservative holists believe that most systems of practices spontaneously respond to the rational end, i.e., the general good. The notion of well-being already appears in the content of some moral requirements: we may for example be required to promote the well-being of others. More widely, it is believed that the systems of practice may display a general tendency towards the general good. Mill and Sidgwick famously argued that our maxims of justice and our common sense morality could be systematised under the idea of the general good. To say this does not mean that agents are generally motivated to act so as to maximise the general good. It is rather to show that whatever motivations agents are typically disposed to act on within a certain system of practice tends, to at least some degree, to realise the general good. The problem, however, is that these systems also respond to other ends corrupting their functioning: egoism, i.e., the pursuit of one's well-being even when in clear conflict with the general good, is one such end. Hence, the need for rational revision or correction.<sup>54</sup> The conservative holistic position I am trying to sketch is nicely summarised in the following quotation from Mill:

---

<sup>54</sup> I am here indebted to Skorupski, 1989, 323-25.

There is no difficulty in proving any ethical standard whatever to work ill, if we suppose universal idiocy to be conjoined with it; but on any hypothesis short of that, mankind must by this time have acquired positive beliefs as to the effects of some actions on their happiness; and the beliefs which have thus come down are the rules of morality for the multitude, and for the philosopher until he has succeeded in finding better. That philosophers might easily do this, even now, on many subjects; that the received code of ethics is by no means of divine right; and that mankind have still much to learn as to the effects of actions on the general happiness I admit, or rather, earnestly maintain.<sup>55</sup>

I would endorse the content of this quotation with the simple substitution of the word ‘good’ instead of Mill’s ‘happiness’. In §7.2., when defending moderate welfarism against charges of over-demandingness and ethical dislocation, the relevance of this discussion of conservative holism will come to full light.

### 6.3. The impartial good, duty, and the extension of ‘morally wrong’

Morality reasons arise independently from reasons concerning the impartial good, yet I am claiming that the impartial good has a bearing in determining the extension of “wrong”. We saw that our moral practices are regulated by a type of rationality internal to the feelings from which they stem. However, when the moral affections enter the realm of action, it is possible to apply a criterion that is external to them. This criterion is the ultimate standard of practical reason, the good impartially considered. If an action or a practice stemming from a particular system of morality is in clear conflict with a principle derived from the reflectively agreed upon impartial good, then there cannot be most reason to pursue that action or practice.<sup>56</sup>

---

<sup>55</sup> Mill, 1998, 70, II24.

<sup>56</sup> This of course implies that considerations of impartiality are always overriding. Someone may want to challenge our position radically. Why should the impartial good have the role of ultimate standard of practical reason and its reasons be always overriding? The justification for this relies on the conclusions reached in Chapter 8. There we argued that when operating under the assumption that our choices will not affect anyone else’s well-being, considerations stemming from the agent’s own well-being enjoy greater rational authority over other kinds of reasons. When that assumption is dropped, it is hard to resist the thought that the kind of consideration enjoying greater rational authority must be some positive function of that which matters most to each individual taken singularly.

If the extension of “morally wrong” were exhaustively determined by its internal criterion, however, a serious problem would arise. Take a case of conflict between a morality reason and an impartial reason, a case in which what we have most reason to do goes *against* a moral obligation. It would be morally wrong to  $\phi$  and yet  $\phi$ -ing brings about the best outcome, i.e., there is most reason to do a morally wrong action. This, I think, would be unacceptable, as it would clash with the claim to categoricity that our moral obligations seem to enjoy. It might be worth running this argument in more detail.

Many believe, and I am inclined to join them, that to say that  $\phi$ -ing is morally wrong conceptually involves either one of two theses; a weaker thesis to the effect that if an action is morally wrong then there is reason not to do it, and a stronger thesis to the effect that if an action is morally wrong then there cannot be most reason to do it. Let us refer to these two theses as to respectively the weak and strong categoricity thesis. In what follows, I will assume that strong categoricity is true. To assert this thesis is to accept that *morality* reasons always override prudential reasons. In order to see why this would be the case one should understand what is involved by the idea of blameworthiness. At this stage, the debate becomes very much a conceptual one and here I shall simply state my inclination to accept the truth of the following thesis: to say that there is reason to blame someone means *inter alia* that considerations of rational self-interest cannot ever *completely* exculpate the blameworthy agent. If in one case they could, that would be sufficient to show that we were not after all in the presence of a blameworthy action.<sup>57</sup>

---

<sup>57</sup> David Brink considers as plausible a view that dissociates the categorical nature of morality (by categorical he here means non-hypothetical) from its rational authority. Morality and the norms it delivers are compared to etiquette and its norms. Its norms are categorical as they apply to agents even when it is not in the agents' interest to abide by them. Morality, however, does *not* deliver categorical *reasons*. Brink believes that there are two senses of 'reason'. Morality reasons, like etiquette reasons, exist “when the relevant sort of behavioral standard or norm” exists. “The other sense of 'reason'... signifies more than the existence of a certain sort of behavioral norm; it indicates that there is a reason to behave in accordance with such a norm such that failure to behave in that way is *ceteris paribus* or *pro tanto* irrational. If there is reason, in this sense, to act on a norm, then practical reason endorses this norm.” Brink, 1992, 8. Brink claims that only prudential reasons are reasons of the latter kind. The idea that morality reasons are norms of the same kind as the norms delivered by etiquette seems to me profoundly mistaken. Norms of etiquette as such, I would claim, provide no justificatory reasons at all. The fact that a norm is socially approved does not in and of itself confer any value to it. An agent can appeal to a certain socially approved norm in order to *explain* his action; that, however, might not at all provide a *justification* for it. Many societies past and present approved of many unjustifiable practices and norms. Social practices may even explain why we pursue some actions that are very detrimental for our own well-being (think about the widespread practice of heavy drinking and smoking in pubs). The goodness of an action for an agent in particular and thus (through the buck-passing view) the reason an agent has to desire and pursue this action, is not derived from or delivered by any of the system of norms Brink has in mind. Just as a system of norms may deliver the wrong prudential dictates, a system of norms may deliver the wrong moral dictates. That I take to be sufficient proof that (socially approved) systems of norms of this kind do not as such provide justificatory reasons for either prudence or morality.

Suppose then that the strong categoricity thesis was true. Its truth would entail that moral reasons enjoy greater rational authority over other types of reasons, such as prudential reasons. That, however, would still leave our initial question unsettled. We would still not be in a position to assert whether the extension of 'morally wrong' can only be determined through the moral feelings or whether, as I would like to suggest, it could also be determined by the impartial good. Compare strong categoricity:

(SC) There can never be most reason to do an action that is morally wrong.

with the idea of the impartial good as the ultimate standard of practical reason:

(MW) There can never be most reason to do an action that does not bring about the best outcome in terms of the impartial good.

Now consider the case in which  $\phi$ -ing has two features: it is morally wrong and it is what brings about the best outcome in terms of the impartial good. According to (SC) there cannot be most reason to  $\phi$ . Yet, if there were only one action that we have most reason to do, according to (MW) it would be false to claim that there cannot be most reason to  $\phi$ , and, in fact, there would be most reason to  $\phi$ . In this case theses (SC) and (MW) would be contradictory. If we did not allow for the possibility that the impartial good determines the content of our moral obligations (if our moral obligations were always exhaustively determined by their internal criterion) then this kind of conflict could always arise; we would have to renounce one of these two theses. These theses are however worth being preserved conjointly. On the one hand, losing (SC) would amount to losing what many would take to be a defining feature of our moral obligations. On the other hand, we would want to have a thesis capable of accommodating supererogation and (MW) is fit to do that. If we could not appeal to a rational criterion outside the morality system such as the

---

This point leaves someone holding the position described by Brink with either of two options. One may deny that moral reasons are justificatory reasons; but then one must come up with a good explanation for the asymmetry between prudential and moral reasons: on what grounds should the former be taken to belong to practical reason and not the latter? After all, as just shown, systems of norms also deliver non-justificatory prudential reasons. Alternatively, one may accept that moral norms are reasons 'endorsed by practical reason' on a par with prudential reasons. Let us simply set aside the former option as too counterintuitive. Now, then, moral reasons and prudential reasons must be considered as at least on a par and this brings home the weak categoricity thesis: if an action is morally wrong, there is reason not to do it.

impartial good, we could not explain how in certain circumstances agents have most reason to do an action that goes beyond the call of duty.

We are thus left with the following picture. When the criterion internal to morality declares  $\phi$ -ing morally wrong and yet there is most reason to  $\phi$  under the idea of the impartial good, then our morality judgement concerning the wrongness of  $\phi$ -ing ought to be revised. It does not follow from this that we are under a moral obligation to do all and only those actions that are considered to be the best under the idea of the impartial good. If  $\phi$ -ing were supererogatory, for example, we may have most reason to  $\phi$  without being morally obligated to do so. It would be morally permissible not to  $\phi$  and to do one's duty instead.

In conclusion, then, given that deontological principles cannot simply be deduced from consideration of the impartial good, moderate welfarism as expressed by (MW) is to be understood as a corrective conservative holistic position. (MW) is still a welfarist position as the ultimate standard of practical reason is the impartial good and this is a positive function of the well-being of all and nothing else.

## **7. Moderate Welfarism: towards a complete defence**

We are now to consider how this version of moderate welfarism fares when confronted with certain thorny issues in ethical theory: supererogation, over-demandingness, and rational egoism. I will provide no further presentation of these issues than that which is offered in the following three subsections.

### **7.1. Supererogatory actions**

The first important issue arising from our formulation of moderate welfarism has to do with supererogation.<sup>58</sup> An agent might have most reason to do something that goes beyond

---

<sup>58</sup> It might be useful, throughout this subsection, to keep this example of supererogation in mind: "Fred is holed up in a defensive position with his fellow soldiers. He is fighting against an evil enemy whose victory would greatly damage the general good. Now a grenade is thrown in. Fred's got three options. His best

the call of duty. More precisely, though he has a moral obligation to act in a certain way, consideration of the general good suggests that he has more reason to act in a supererogatory way. It is crucial to bear in mind that cases of supererogation do not involve moral conflicts. Supererogatory actions go *beyond* duty, not *against* it. Unlike the case discussed in the last section, then, whatever the agent chooses to do is morally permissible. The agent may either simply attend to the call of duty or do the supererogatory act: either way he would not be blameworthy. The agent, however, has most reason to do the supererogatory action, even though he is not under a moral obligation to do it. In the case in which the agent decided to stick to his moral duty, he is less than optimally rational, as he does not do what he has most reason to do. We shall get back to this issue presently.

First, however, the following should be noticed. Often the agent who goes beyond the call of duty is a lover of virtue in the sense that he takes himself to have reason to *desire* to do the supererogatory action. Alternatively, he might rightly envisage the supererogatory act as an achievement, and have reason to *desire* to pursue it as such. It is perfectly fine to let well-being reasons play an active role at this level. Taking both the personal and the impartial point of view into account, the agent would consider the supererogatory action to be more choiceworthy. This, however, does not exclude the case in which an agent can clearly see that he has reason not to *desire* the supererogatory action but simply reason to pursue it, and yet does not pursue it as he is not morally obliged to do so.

This takes us to our second remark. Presumably, what morality requires in a given society at a given time may not always be as determinate and clear-cut as we might want; this is mostly so in societies where moral practices are commonly submitted to rational validation and scrutiny. The same goes for what is required by considerations stemming from the impartial good. Certain agents may have good reason to believe that what duty requires one to do on a certain occasion is not quite what they would see as optimal to do. Other

---

chance of saving himself is to run away, his next best is to take evasive action by throwing himself to the ground. But he could also throw himself on to the grenade. This last action, he knows, would save the most people. For he knows that his comrades won't run away—they can see that they have an obligation to stay at the post and go on fighting, and they'll do that. But they won't throw themselves on the grenade either. So if he throws himself on it, stifling its impact, more soldiers will be saved and the chance of resisting the evil enemy will be greater. This, then, is the action that will most promote general good. Coming to this conclusion, he throws himself on the grenade. ... the right *moral* assessment is that Fred went beyond the call

agents, however, may disagree and may do so in two ways: they may disagree either about what duty requires or about what the impartial good requires.<sup>59</sup> Ideally, the two parties would have to engage in rational dialogue concerning that practice. But even then, agreement is not guaranteed. If even the reflected opinion concerning certain practices is not in agreement, different individuals or groups of individuals might be warranted, at least temporarily, in holding contrasting beliefs as to what there is most reason to do. Importantly, for our initial problem, it would be misplaced to state in these kinds of cases that there is most reason to do one rather than the other action and/or practice. Agents who can't see why there would be most reason to do the supererogatory action cannot be considered as less than optimally rational when doing 'only' what they are morally required to do.

## 7.2. Demandingness and dislocation

The corrective view may be *wrongly* considered to be a version of the view that the general good provides the justification for all moral reasons: for example, a principle of action is morally wrong to the degree to which it fails to bring about the most general good. In the last thirty years or so, consequentialist moral theorists have been pressed with the charge that the impartiality required by their theories was too demanding on the agents who tried to put it in practice in their daily lives. Agents who seriously attempted to follow its prescriptions would become alienated and schizophrenic, incapable of pursuing their own projects and aims or committing to special relationships with their friends and loved ones without a sense of culpability or strain. In the words of Williams, these theories often involved "one thought too many" in the motivation of the agent. That I should save my wife rather than a stranger from drowning because it is *my* wife is all I should be motivated by when choosing to save her rather than the stranger. But according to Williams, this is not what these theories prescribe.

---

of duty. Fred and his comrades had an obligation not to run away; but they did not have an obligation to fall on the grenade. Our morality is not heroic..." Skorupski, 2001, 72-73.

<sup>59</sup> It would be interesting to see how present Western opinions concerning abortion would fit this way of setting out the problem.

As a response to these charges, some consequentialists (and even some deontologists) have adopted the strategy of separating two levels of thinking or inquiry: the 'intuitive' level (or decision-theoretic) and the 'critical' level (or as it sometimes referred to, the 'standard of rightness').<sup>60</sup> Though their principle of impartiality (be it Kantian or Utilitarian) had a justificatory role, it needn't appear at the level of our every day motivation.<sup>61</sup> This kind of move was already familiar through J. S. Mill's and Sidgwick's forms of multi-level or indirect utilitarianism. Williams, however, charges this type of defence with what we may call the dislocation problem.

### Williams's charge

It might be worth quoting Williams at length on this issue:

These styles of indirect utilitarianism involve a special view of the dispositions that are exercised at the everyday or intuitive level; and this raises a serious question: Is there anywhere in the mind or in the society that a theory of this kind can be coherently or acceptably located? The theory finds a value for these dispositions, but it is still an instrumental value. The dispositions are seen as devices for generating certain actions, and those actions are the means by which certain states of affairs, yielding the most welfare, come about. This is what dispositions look like when seen from outside, from the point of view of the utilitarian consciousness. But it is not what they seem from the inside. Indeed, the utilitarian argument implies that they should *not* seem like that from the inside. The dispositions help to form the character of an agent who has them, and they will do the job the theory has given them only if the agent does not see his character purely instrumentally, but sees the world from the point of view of that character. Moreover, the dispositions require the agent to see other things in a noninstrumental way. They are dispositions not

---

<sup>60</sup> See Stark, 1997, for a defence of deontological theories from the charge over-demandingness.

<sup>61</sup> Ashford, 2003, offers another reply to the charge of over-demandingness. Having shown that even Kantian contractualist accounts such as Scanlon's, 1998, Chapters 4-5, fall prey to this charge, Ashford suggests that "The extreme demandingness of both contractualist and utilitarian obligations to those in need is not an objection to either view... but an appropriate response to morally salient features of the current state of the world. In our world, we would expect any theory grounded on the claim that each person has equal moral status to hold that agents have extremely demanding obligations to give aid... It can plausibly be claimed that both utilitarian and Kantian contractualist obligations to those in need would be considerably less demanding if the state of the world were relevantly different." Ashford, 2003, 292-293. Or again: "...if we take seriously the central tenet of enlightenment moral theory that each person has equal moral status, we may have to accept that our current moral obligations to those in need are drastically more demanding than our commonsense moral thinking tells us." Ashford, 2003, 275.



only of action, but of feeling and judgment, and they are expressed precisely in ascribing intrinsic and not instrumental value to such things as truth-telling, loyalty and so on.

There is a deeply uneasy gap or dislocation in this type of theory, between the spirit of the theory itself and the spirit it supposedly justifies. There is a distinction that is supposed to bridge the gap or, rather, make us accept it: the distinction between theory and practice. But when one asks whose theory is in question, and whose practice, the distinction turns out to have very little power.<sup>62</sup>

To show his point about the problems afflicting this distinction, Williams looks at the way both Sidgwick and Richard Hare put it forward. The former expounds a version of indirect utilitarianism Williams refers to as "Government House utilitarianism." The idea is that enlightened utilitarians might be able to live by rules that should not be divulged to the masses, as that would be likely to produce more harm than good by weakening morality. According to Sidgwick, it would also follow from this view that, for example, it may be right to do and privately recommend what it would not be right to advocate openly. Or again, it may be right to teach openly to one set of persons what it would be wrong to teach to others.<sup>63</sup> In Hare, one finds that the distinction between theory and practice is cashed out in psychological rather than social terms. There is the time for theorising and the time for practice. Williams, finds that:

It is difficult to suppose that a thorough commitment to the values of friendship and so on can merely alternate, on a timetable prescribed by calm and activity, with an alien set of reflections. Moreover since the reflections are indeed alien, some kind of willed forgetting is needed, an internal surrogate of those class barriers on which Sidgwick relied, to keep the committed dispositions from being unnerved by instrumental reflection when they are under pressure.<sup>64</sup>

Williams draws the conclusion that the kind of life the utilitarian agent would be required to live is a dislocated one,<sup>65</sup> one with an unacceptable gap. In Sidgwick's case, the gap is social: the agent has to treat his fellow human beings according to their utilitarian intellectual enlightenment. Regarding Hare, the gap is a personal one and, I would add,

---

<sup>62</sup> Williams, 1985, 107-108.

<sup>63</sup> See Sidgwick, 1907, 489-490.

<sup>64</sup> Williams, 1985, 109.

<sup>65</sup> Williams writes: "The discussion in the present chapter concerns what kind of life, social or personal, would be needed to embody such a theory." 1985, n. 12, 215.

somehow more radical: the agent's theoretical reflections about his actions cannot systematically be brought to bear on the practice of his actions. Some sort of an invisible moral hand ensures that the agent's typical motivation and/or dispositions in action are of the right kind.

#### A reply from the conservative holist

Our formulation of moderate welfarism acknowledges the force of Williams's point. We laid so much stress on the fact that our well-being reasons and our morality reasons arise independently from the general good. The general good, however, functions mainly as a constraint. We are not *deriving* our reasons for action, morally and otherwise, from it. Individuals are situated in a determined social-historical context. What they take themselves to have reason to feel and do is partly determined by their socio-historical context. Conservative holism, takes stock of that, and proposes only to *correct* the practices embodied in this context when they conflict with the general good.

Does this position involve levels of thinking? Would it be an alienating, or over-demanding position? Or would it prove to involve an unacceptable gap in the agent's psychological outlook? In Sidgwickian spirit, I would like to distinguish between two levels of thinking: that of the agent's day-to-day moral deliberations and that of the moral theorist, or again, the level of ethical life and the level of ethical theory. As we shall see, however, this distinction is nowhere as clear-cut as Sidgwick's nor does it carry the problems pointed out by Williams. Our (ethical) life may confront us with choices that involve going beyond what one is morally required to do. We examined these cases under the heading of supererogation. We also saw that agents may have reason to desire to choose the supererogatory action. What is more, to allow agents to make a direct appeal to consideration of the general good in their everyday lives is to allow the possibility of experiments in individuality: some agents may have reason to desire to live supererogatory lives as a way of realising their autonomy.

It is important to notice, however, that our view does not impose this kind of act upon everyone. One should remember that going beyond the call of duty *must* be morally

permissible and not morally required. Also, it may very well be that what makes a supererogatory action the best impartial outcome is precisely the fact that it would be an isolated instance.<sup>66</sup> It is not implausible to think that if all of the agents in a society were required to go beyond the call of duty at every time, as a result the general good would suffer rather than be enhanced. Yet, though we do not have most reason to go beyond the call of duty at all times, it is not implausible to think that all of us may at any time face choices in which there is most reason to go *beyond* duty and even *against* it.

The question arises, however, as to who is to determine what actions would bring about the best outcome under the idea of the general good. It is at this stage that ethical theory and ethical theorists are needed. We mentioned already in §2.2. that determining even just the form of the positive function of the well-being of all individuals involves reflective discussion as well as contributions from specialists working within the human and social sciences. Is there anything wrong with this? I do not think so. Just as we might rely on the specialised knowledge of medical doctors to tell us what things to do in order to conduct a healthy life and thus preserve our well-being, we might rely on the specialised knowledge of philosophers, sociologists, economists, etc. in order to get some guidance about what the best possible arrangement for all might be.

Though I have appealed to Sidgwick in order to draw the distinction between ethical life and ethical theory, it should be noticed that drawing this distinction does not necessarily involve the type of 'Government House' utilitarianism attributed to Sidgwick. The specialists are first and foremost individuals situated in a determinate socio-historical context like all other individuals. Their ethical lives will be similar to all others. To be specialists in ethical issues makes them ethicists and being an ethicist does not at all involve having access to some secretive truth that should not be divulged. Unlike Sidgwick's enlightened utilitarians, ethicists are not in the business of *deriving* the right from the good but only in that of determining the impartial good and how, if at all, it conflicts with morality. Maybe they may find out before others that certain practices are not in the interest of all; they may in that case decide to adapt their ways of life accordingly; they might publish papers to divulge their findings; they might lobby the government to issue some new directives. The same goes even if they discovered that

---

<sup>66</sup> Think back to the example of Fred and the grenade. If Fred and, say, Jack both threw themselves on the

certain practices would bring about greater general good only if a minority of the population were allowed to live by them. It must also be remembered, however, that contrary to Sidgwick our conception of individual well-being includes both autonomy and knowledge as intrinsic ends. To keep certain facts hidden from the rest of the public would seem to conflict with both ends. It is not excluded in principle that such specialists might discover facts that are best kept away from public knowledge. But given the normative importance of these two ends for each agent's well-being it seems unlikely in practice that a case might arise in which infringing these two values would be compatible with the best outcome under the idea of the impartial good.

Echoing Williams's words from the long quotation above, I will conclude as follows. Our conservative holistic version of moderate welfarism does not dislocate agents' ethical life. Agents can, from their everyday level of thinking, keep ascribing intrinsic and not instrumental value to such things as truth-telling, loyalty and so on. On some occasions, however, an agent may be confronted with a real life situation in which he takes himself to have most reason to do something that goes beyond or even against the call of duty. The agent may realise that there is a constraint not to do what does not bring about the best outcome under the idea of the general good. Sometimes what is involved by this idea may seem to be straightforwardly clear. At other times it is not. An agent, then, may rely on the informed opinion of those who have spent time thinking about these issues. What is more, agents may differ in their capacities to recognise and abide by the constraints set by the impartial good. Though it is implausible to think that the impartial good could set constraints on all of us at all times it would be a bit too convenient to think the opposite. What balance to strike between these two extremes should pose itself as a difficult question from the point of view of ethical theory as well as from the point of view of any reflective agent.

### 7.3. Why should moderate welfarism be impartial?

---

grenade the result would probably be sub-optimal and not what there would be most reason to do.

In Chapter 8, I discussed moderate welfarism under the assumption that the agent's choice would not affect third parties. Under that assumption, we attempted to defend welfarism as thesis (EMW):

(EMW) There can never be most reason for an agent to do something that doesn't bring about the best outcome in terms of his or her good.

If this same thesis were thought to hold even after having dropped the assumption we just mentioned, then (EMW) would amount to what we may call a moderate version of rational egoism. That's because even though (EMW) could accept the possibility of morality and impartiality reasons for action, it would disallow the possibility of there being most reason to do these actions when in conflict with one's well-being.

Rational egoism, however, is undermined by the distinction between what we called virtuous moral agents and conscientious moral agents. If rational egoism were true, we could not accommodate the possibility of acting conscientiously. It would never be the case that an agent has most reason to act morally if that is incompatible with his well-being. And, yet, many times that seems precisely to be what happens: we act believing that there is 'nothing in it' for us; in these cases it is out of a sense of duty that we choose to act morally. Rational egoism would have to explain away this phenomenon of our ethical life, our sense of duty, as based on an error. It would have to show that rational self-sacrifice is a chimera. (MW), on the other hand, can allow for both the possibility of rational self-sacrifice and that of conscientious action. This is a reason to prefer it to (EMW) as a better version of moderate welfarism.

More positively, I believe pure impartialism can be defended through an argument we have already presented. Realisation on behalf of an agent that well-being has agent-relative value may take the agent one step further towards the realisation that every agent capable of well-being has reason not to desire that any action involving a negative impact on his or her well-being be performed. Next, the agent will have to grasp what bearing this realisation has on his actions. Some agents may simply not be able to recognise that anything follows concerning their actions. I would say that this type of agent does not have the capacity to ever be motivated by the thought that an action having a negative impact on another person's well-being gives reason to consider whether it is appropriate to do it or

not. This type of thinker, however, must accept that others have no reason to have any regard for their well-being, as to think so would make them irrational.<sup>67</sup> Something else must be said about the possible deficiencies in the psychological/cognitive outlook of these agents.

Nagel sees a parallel between intertemporal and interpersonal distribution of harms and benefits. He argues for agent-neutrality by analogy with prudence. Prudence requires that I act now for something that is in my future interest. Failure to recognise prudence involves temporal dissociation, i.e., failure to see the present as just one time among others. Similarly, agent-neutrality requires one to take into account other people's interests. Failure to recognise them involves personal dissociation, i.e., failure to recognise oneself as just one person among others.<sup>68</sup> Failure to be motivated by agent-neutral judgements amounts to a form of practical solipsism, i.e., "the inability to draw fully fledged practical conclusions about impersonally viewed situations."<sup>69</sup> I can see why *I* should be motivated to remove *my* foot when a piece of furniture is about to land on it. However, I am unable to be motivated to avoid dropping the same piece of furniture on *someone else's* foot even though I can, merely in virtue of the fact that it is *his* foot and that may harm *him*. Such a dissociated agent is unable to cognise the principle that other people are real and so are their interests. This agent could still act altruistically through internal motivating factors such as self-interest, love, and sympathy. The principle of the reality of other persons, however, is a formal principle of our rationality capable of motivating our actions on its own. Agents lacking it, we shall conclude, will be cognitively or rationally deficient to some extent.

Nagel's argument, it seems to me, does not give enough weight to the separateness of persons. To discuss that thesis, however, would lead us astray and I would rather emphasise my agreement with the core of Nagel's argument. Grasping the agent-relativity and universalisability of well-being reasons should *ipso facto* point any intellect towards recognising reasons not to harm other people's interests; it should *ipso facto* point at their

---

<sup>67</sup> Their position would amount to that labelled as (IE) or irrational egoism above. Everyone has reason to promote *my* well-being, because it is *my* well-being (but I have no reason to promote anyone else's but my well-being).

<sup>68</sup> This type of argument can be traced back to at least Sidgwick. See for example Sidgwick, 1907, 418-19. More recently it has been greatly elaborated by Parfit, 1984, esp. Part II.

<sup>69</sup> Nagel, 1971, 114.

agent-neutrality. What should we say, however, about those who cannot grasp this primitive link and thus fail to ever be motivated to act accordingly? I believe their cognitive incapacity is serious. In fact, it is so serious that these agents cannot be held responsible for their non-compliance to the dictates stemming from these reasons. I do not believe, however, that there are many genuine cases of these. For all the other thinkers it would be sufficient to grasp that every agent has reason to desire his or her well-being, to also grasp that each agent's well-being must be given equal or impartial consideration. Understanding that well-being has special value will in most cases direct the thinker's intellect towards well-being's impartial value. To understand the impartial value of well-being is to understand that one's well-being is as valuable as anyone else's and typically be motivated to act accordingly. To understand that, however, does not imply that one will unfailingly be motivated to act accordingly. A position that *fully* recognises the impartiality of well-being will no longer be compatible with (EMW).

#### 7.4. Concerns from the deontologist camp

We have begun to defend moderate welfarism from objections coming from two camps: rational egoism and pure teleology. We did not say much, however, about pure deontology and shall not try to argue against it here. I shall, however, acknowledge some of the concerns that pure deontologists may take to affect our account. More precisely, I am thinking of the discussion of Nozick and side-constraints as we broached it in §5.1. Does even our indirect and corrective moderate welfarist view allow that it is possible, at least in principle, to sacrifice the good of someone for the sake of others? Wouldn't that be morally impermissible, as it would conceive of people as of expendable resources? If our view allowed that, then I am inclined to think that there would be some irreparable tension in it.

Whether the impartial good would in principle ever allow the infringement of rights as side-constraints would depend on what distributive principle is reflectively agreed upon. If the principle is aggregate maximization à la Mill, then we could grant the deontologist his worry: there would be an inescapable tension, at least in principle, between the impartial good and some of our moral practices. In principle, aggregate utility can always be increased by systematically sacrificing the well-being of an individual or of a class of individuals. Our fundamental moral attitudes tell us that that is unfair just as is it is unfair to punish someone who did not deserve it. But aggregate maximization is not the only

distributive principle. It is the task of ethical theory to explore the ways in which other distributive principles could accommodate this fundamental tension.

## 8. Conclusion

To defend moderate welfarism along the lines we have sketched above is to defend the thesis that well-being still occupies a central role in ethics. Of course, this version of welfarism is not as strong as that which some welfarists would want to defend. In this as well as the last chapter we did not exclude the possibility of non-welfarist reasons, i.e., reasons for actions unrelated to the well-being of each person taken agent-relatively or agent-neutrally. Abstracting from moral and impartial considerations, however, the reasons grounded on a person's own well-being seem to enjoy greater rational authority. Thus, even though the possibility of other kinds of reasons for action is not excluded, their reason-giving force seems to be checked by reasons concerning one's well-being. Similarly, when moral and impartial considerations are brought back into the picture. The existence of independent morality reasons is undeniable, or at least so I have argued. What these reasons enjoin us to do, however, can be corrected. Reasons stemming from the impartial good enjoy greater rational authority over both morality reasons and reasons stemming from consideration of one's own well-being. This should be read as a statement of (MW).

My version of moderate welfarism encounters opposition from at least three different camps: rational egoism, pure deontology, and pure teleology. I spent the most time defending moderate welfarism against pure teleology, by showing the implausibility of deriving the right from the good. I also sketched a brief defence against rational egoism and Williams's charge might also be taken to share some of the worries of deontology (though surely not from a point of view such as Kant's). However, some of the deontologists' most fundamental worries could not be fully addressed in this thesis. Concerning the tension between the impartial good and rights as side-constraints we could only point at the irreparable tension between our moral attitudes on the one side and aggregate maximization of well-being on the other. I suggested that someone wanting to solve this tension might only have the option of exploring the possibilities offered by



different distributive principles, ones that can be brought closer to some of our most fundamental moral attitudes.

## Conclusion

In this thesis I have described the role of well-being in ethics. In Part I, I argued that the concept of well-being does play an explanatorily important role in the thinking of a rational agent. I arrived at this conclusion by distinguishing levels of thinking activity as well as by considering the implicit rather than explicit role well-being plays in our deliberation. *Contra* Scanlon, I also claimed that the concept of well-being is not necessarily a concept shaped by moral assumptions (at least on his definition of what it is for something to be moral). We concluded this part of the thesis by illustrating the relation between the idea of well-being, its parts and its sources. Well-being is an inclusive good. The idea of well-being is an insubstantial regulatory ideal. An agent has reason to desire the various parts of well-being for their own sakes (where that doesn't mean that we have reason to desire them irrespective of context) under the idea of his good or well-being. If on one occasion it could be shown that it would be detrimental for me to desire an instance of enjoyment it cannot be most rational for me to desire it, unless considerations other than my own well-being are also in play. I believe that Scanlon has gone too far in playing down the explanatory importance of the concept of well-being. I take Butler's picture to be much closer to our intuitions at least to the extent that it firmly established the explanatory role of self-love as a principle of action. My own picture has the advantage of assigning to well-being the explanatory role Butler understood so well.

According to Scanlon, we can at most offer "a theory of what is valuable in general" rather than a unified theory *of well-being*. That's because, on his account, a person's well-being is largely contributed to by achieving success in his rational aims and not all of a person's rational aims are chosen for their contribution (nor do they contribute) to that person's well-being. In Part II, I put forward a unified theory *of well-being*. I did so firstly by taking on board Scanlon's own buck-passing account of value and, with some help from Mill, by arguing that the 'good for a person' is what there is reason for this person to desire. To say this, however, does not exclude that an agent might have reason to pursue many things he does not have reason to *desire*. Secondly, I showed that Scanlon's rational aim account should be reduced to the reason-to-desire account. Thirdly, I claimed that a person's life worthiness shouldn't be confused with this person's well-being. Sure enough, many

individuals take themselves to have reason to *pursue* more admirable lives. That, however, does not imply that there is reason for each of us to *desire* this type of life for oneself. Fourthly, I defined the notion of rational self-sacrifice, I illustrated how our account of well-being is not affected by the ‘scope problem’, and I showed how Scanlon seems to be using the notion of self-sacrifice inconsistently. Much of this part of the thesis was dedicated to a number of different though connected projects: the defence *contra* Darwall of the claim that well-being is itself a normative notion; the description of the expansion of the self as a defence of the claim that we have reason to desire all and only things that are parts of our good. Finally, I raised various specific issues that concern the particular components of well-being.

Having determined the explanatory role of well-being from a first-personal point of view, and having offered a systematic account of well-being, in Part III, I began to sketch the normative role of well-being both first-personally and impartially. With Scanlon, I agreed that well-being cannot be a “master value”. Yet, I argued in favour of a moderate form of welfarism, a view that takes a positive function of each individual’s well-being to afford the ultimate criterion of practical reason. The last two chapters do not afford a complete defence of moderate welfarism. I would however consider the ideas I formulate in these chapters as providing the framework within which further research in normative ethics should be pursued.

## Bibliography

- ACKRILL, J.L. (1980) "Aristotle on *Eudaimonia*." In A. Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics*. Berkeley and Los Angeles: University of California Press.
- ANNAS, J. (1994) "Response to Crisp." In *Philosophical Books*, 34.  
 (1995) "Prudence and Morality in Ancient and Modern Ethics." In *Ethics*, 105.1.
- ARISTOTLE (1984) *Nicomachean Ethics*. Trans. W.D. Ross. In Barnes J. (ed.), *The Complete Works of Aristotle*, 2 vols. Princeton: Princeton University Press.
- ARNESON, R. (1999) "Human Flourishing Versus Desire Satisfaction." In *Social Philosophy and Policy*, 16.1.
- ASHFORD, L. (2003) "The Demandingness of Scanlon's Contractualism." In *Ethics*, 113.2.
- BARON M. (1991) "Impartiality and Friendship." In *Ethics*, 101.4.
- BERGER, F. (1984) *Happiness, Justice, and Freedom*. London: University of California Press.
- BRANDT, R. (1979) *A Theory of the Good and the Right*. Oxford: Clarendon.
- BRINK, D. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.  
 (1992) "A Puzzle About the Rational Authority of Morality." In *Philosophical Perspectives*, 6.
- BROAD, C.D. (1930) *Five Types of Ethical Theory*. London: Kegan Paul.
- BROADIE, S. (1991) *Ethics with Aristotle*. Oxford: Oxford University Press.
- BROOME, J. (1991) *Weighing Goods*. Oxford: Basil Blackwell.  
 (1999) *Ethics out of Economics*. Cambridge: Cambridge University Press.
- BUTLER, J. (1949) *Fifteen Sermons Preached at the Rolls Chapel*. Revd. D. Matthews (ed.). London: Bell & Sons.
- CHRISTIANO, T. (1992) "Sidgwick on desire, pleasure, and the good." In B. Schultz (ed.) *Essays on Sidgwick*. Cambridge: Cambridge University Press.
- CLIFFORD, W. (1879) *Lectures and Essays*, Vol. II. F. Pollock (ed.). London: Macmillan and Co.
- COTTINGHAM, J. (1991) "The Ethics of Self-Concern." In *Ethics*, 101.
- CRISP, R. (1990) "Sidgwick on Self-Interest." In *Utilitas*, 2.2.  
 (1992) "White on Aristotelian Happiness." In *Oxford Studies in Ancient Philosophy*, 10.  
 (1994a) "Values, Reasons, and The Environment." In R. Attfield & A. Belsey (eds.), *Philosophy and The Natural Environment*. Royal Institute of Philosophy, Supplement 36.  
 (1994b) "Aristotle's Inclusivism." In *Oxford Studies in Ancient Philosophy*, 12.  
 (1994c) "Ancient Wisdoms." In *Philosophical Books*, 34.  
 (1997a) *Mill on Utilitarianism*. London: Routledge.  
 (1997b) "Raz on Well-Being." In *Oxford Journal of Legal Studies*, 17.3.  
 (2000a) "Value... and What Follows. By Joel Kupperman." In *Philosophy*, 75.  
 (2000b) "Utilitarianism and accomplishment." In *Analysis*, 60.3.  
 (2001) "Utilitarianism and accomplishment revisited." In *Analysis*, 61.2.
- CRISP, R. & HOOKER, B. (eds.) (2000) *Well-Being and Morality: Essays in Honour of James Griffin*. New York: Oxford University Press.
- CRISP R. & MOORE A. (1996) "Welfarism in Moral Theory." In *Australasian Journal of Philosophy*, 74.4.
- DANCY, J. (2000) "Should we pass the buck?" In *Philosophy, the Good, the True and the Beautiful*, Cambridge: Cambridge University Press.
- DARWALL, S. (1977) "Two Kinds of Respect." In *Ethics*, 88.  
 (1992) "Internalism and Agency." In *Philosophical Perspectives*, 6.  
 (1997) "Self-Interest and Self-Concern." In *Social Philosophy and Policy*, 14.1.  
 (2000) "Sidgwick, Concern, and the Good." In *Utilitas*, 12.3.  
 (2002a) "John Skorupski, *Ethical Explorations*." In *Utilitas*, 14.1.

- (2002b) *Welfare and Rational Care*. Princeton: Princeton University Press.
- (2003) "Moore, Normativity, and Intrinsic Value." In *Ethics*, 113.3.
- DUNCAN-JONES, A. (1952) *Butler's Moral Philosophy*. London: Penguin Books.
- DWORKIN, G. (1986) *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- ENGSTROM S. (1996) "Happiness and the Highest Good in Aristotle and Kant." In S. Engstrom and J. Whiting (eds.), *Aristotle, Kant, and the Stoics*. New York: Cambridge University Press.
- FEINBERG, J. (1986) *Harm to Self*. Oxford: Oxford University Press.
- FINNIS, J. (1980) *Natural Law and Natural Rights*. Oxford: Clarendon Press.
- FRANKENA, W. (1993) "Sidgwick's *Methods of Ethics*, Edition 7, Page 92, Note 1." In J. Heil (ed.), *Rationality, Morality, and Self-Interest*. Maryland: Rowman and Littlefield Publishers.
- FRANKFURT, H. (1988) *The Importance of What We Care About*. Cambridge: Cambridge University Press.
- FREY, R.G. (1992) "Butler on Self-love and Benevolence." In C. Cunliffe (ed.), *Joseph Butler's Moral and Religious Thought*. Oxford: Clarendon Press.
- FUCHS, A. (1993) "Posthumous Satisfaction and the Concept of Individual Welfare." In J. Heil (ed.), *Rationality, Morality, and Self-Interest*. Maryland: Rowman and Littlefield Publishers.
- GRIFFIN, J. (1986) *Well-being: Its Meaning, Measurement and Moral Importance*. Oxford: Oxford Clarendon Press.
- (1988) "Happiness." In E. Craig (ed.) *Routledge Encyclopaedia of Philosophy*, Vol. 4. London: Routledge.
- (1996) *Value Judgement: Improving Our Ethical Beliefs*. Oxford: Oxford University Press.
- (2000) "Replies." In R. Crisp & B. Hooker (ed.), *Well-Being and Morality: Essays in Honour of James Griffin*. New York: Oxford.
- HAYBRON, D. (2001) "Happiness and Pleasure." In *Philosophy and Phenomenological Research*, 62.3.
- HENSON, R. (1988) "Butler on Selfishness and Self-Love." In *Philosophy and Phenomenological Research*, 49.1.
- HILL, T. JR. (2002) *Human Welfare and Moral Worth*. Oxford: Oxford University Press.
- HOOKE, B. (2000) "Moral Particularism: Wrong and Bad." In B. Hooker and M. O. Little (eds.), *Moral Particularism*. Oxford: Clarendon Press.
- HUBIN, D. (1996) "Hypothetical Motivation." In *Nous*, 30.1.
- HUME, D. (1978) *A Treatise of Human Nature*. L.A. Selby-Bigge (ed.), 2<sup>nd</sup> ed. Hong Kong: Oxford University Press.
- HURKA, T. (1993) *Perfectionism*. New York: Oxford University Press.
- (1999) "Three Faces of Flourishing." In *Social Philosophy and Policy*, 16.1.
- IRWIN, T.H. (1977) *Plato's Moral Theory*. Oxford: Clarendon Press.
- (1995) "Prudence and Morality in Greek Ethics." In *Ethics*, 105.1.
- (1996) "Kant's Criticisms of Eudaimonism." In S. Engstrom and J. Whiting (eds.), *Aristotle, Kant, and the Stoics*, New York: Cambridge University Press.
- KAGAN, S. (1992) "The Limits of Well-Being." In *Social Philosophy and Policy*, 9.2.
- KANT, I. (1968) *Critique of Pure Reason*. Trans. N. Kemp Smith. London: Macmillan.
- (1996) *Practical Philosophy*. Trans. M. J. Gregor (ed.). Cambridge: Cambridge University Press.
- KORSGAARD, C. (1996) *Creating the Kingdom of Ends*. Cambridge New York: Cambridge University Press.
- LOEB, D. (1995) "Full-Information Theories of Individual Good." In *Social Theory and Practice*, 21.
- MACKIE, J.L. (1977) *Ethics, Inventing Right and Wrong*. London: Penguin.
- (1992) "Sidgwick's pessimism." In B. Schultz (ed.) *Essays on Sidgwick*. Cambridge: Cambridge University Press.
- MCDOWELL, J. (1980) "The Role of *Eudaimonia* in Aristotle's Ethics." In A. Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics*. Berkeley and Los Angeles: University of California Press.

- (1996) "Deliberation and Moral Development in Aristotle's Ethics." In S. Engstrom and J. Whiting (eds.), *Aristotle, Kant, and the Stoics*. New York: Cambridge University Press.
- MCNAUGHTON, D. (1992) "Butler on Benevolence." In C. Christopher (ed.), *Joseph Butler's Moral and Religious Thought*. Oxford: Clarendon Press.
- MCNAUGHTON, D. & RAWLING, P. (2001) "Achievement, welfare, and consequentialism." In *Analysis*, 61.2.
- MILL, J.S. (1936) *A System of Logic*. London: Routledge & Kegan Paul.
- (1977) "On Liberty." In J. M. Robson (ed.) *Collected Works of John Stuart Mill, Volume XVIII, Essays on Politics and Society*. Toronto, London: University of Toronto Press, Routledge.
- (1998) *Utilitarianism*. Roger Crisp (ed.). Oxford: Oxford University Press.
- MOORE, A. (1991) *A Theory of Well-Being*. (D.Phil. Thesis), University of Oxford.
- MOORE, G.E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- (1912) *Ethics*. Oxford: Oxford University Press.
- (1922) *Philosophical Studies*. London: Kegan Paul.
- NAGEL, T. (1978) *The Possibility of Altruism*. Princeton: Princeton University Press.
- (1986) *The View From Nowhere*. New York: Oxford University Press.
- NOZICK, R. (1974) *Anarchy, State, and Utopia*. Oxford: Basil Blackwell.
- (1989) *The Examined Life*. New York: Simon & Schuster.
- OVERVOLD, M. (1980) "Self-interest and the Concept of Self-sacrifice." In *Canadian Journal of Philosophy*, 10.1.
- (1982) "Self-Interest and Getting What You Want." In H. Miller and W. Williams (eds.), *The Limits of Utilitarianism*. Minneapolis: University of Minnesota Press.
- (1993) "The Relation of Pleasure to Desire in Butler's Moral Psychology." In J. Heil (ed.), *Rationality, Morality, and Self-Interest*. Maryland: Rowman and Littlefield Publishers.
- PARFIT, D. (1984) *Reasons and Persons*. Oxford: Clarendon Press.
- (2001) "Equality or Priority?" in J. Harris (ed.) *Bioethics*. Oxford: Oxford University Press.
- PENELHUM, T. (1985) *Butler*. London: Routledge & Kegan Paul.
- PHILLIPS, D. (2000) "Butler and the Nature of Self-Interest." In *Philosophy and Phenomenological Research*, 60.2.
- PITCHER G. (1984) "The Misfortunes of the Dead." *American Philosophical Quarterly*, 21.2.
- QUINN, W. (1993) *Morality and Action*. Cambridge: Cambridge University Press.
- RABINOWICZ, W. & RØNNOW-RASMUSSEN T. (unpublished) "The Strike of the Demon: On Fitting Pro-Attitudes and Value."
- RAILTON P. (1986) "Facts and Values." In *Philosophical Topics*, 14.2.
- RAWLS, J., (1972) *A Theory of Justice*. Oxford: Clarendon Press.
- RAZ, J. (1984) *The Morality of Freedom*. Oxford: Clarendon Press.
- (2000) "The Central Conflict: Morality and Self-Interest." In R. Crisp and B. Hooker (eds.), *Well-Being and Morality: Essays in Honour of James Griffin*. New York: Oxford University Press.
- REGAN, D. (1989) "Authority and Value: Reflections on Raz's *Morality of Freedom*." In *Southern California Law Review*, 62.
- ROSATI, C. (1995) "Person, Perspectives, and Full Information Accounts of the Good." In *Ethics*, 105.1.
- RUBEN, D. (1989) "A Puzzle About Posthumous Predication." In *Philosophical Review*, 97.
- SAYRE-MCCORD, G. (2001) "Mill's "Proof" of the Principle of Utility: a More than Half-hearted Defense." In *Social Philosophy and Policy*, 18.2.
- SCANLON, T. (1991) "The moral basis of interpersonal comparisons." In J. Elster and J. E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*. New York: Cambridge University Press.
- (1998) *What We Owe to Each Other*. Cambridge, Mass., London: Belknap Press of Harvard University Press.
- (2001) "Sen and Consequentialism." In *Economics and Philosophy*, 17.1.
- SCARRE, G. (1999) "Happiness for the Millian." In *British Journal for the History of Philosophy*, 7.3.

- (2001) "On Caring About One's Posthumous Reputation." *American Philosophical Quarterly*, 38.2.
- SCHEFFLER, S. (1982) *The Rejection of Consequentialism*. Oxford: Clarendon Press.
- SCHNEEWIND, J.B. (1977) *Sidgwick's Ethics and Victorian Moral Philosophy*. New York: Clarendon Press.
- SEN, A. (1982) "Rights and Agency." In *Philosophy and Public Affairs*, 11.  
 (1985a) "Well-being, Agency, and Freedom." In *The Journal of Philosophy*, 82.1.  
 (1992) *Inequality Reexamined*. Cambridge Mass.: Harvard University Press.  
 (1993) "Capability and Well-Being." In M. Nussbaum and A. Sen (eds.), *The Quality of Life*. Oxford: Oxford University Press.  
 (2000) "Consequential evaluation and practical reason." In *The Journal of Philosophy*, 96.  
 (2001) "Reply." In *Economics and Philosophy*, 17.1.
- SHAVER, R. (1997) "Sidgwick's False Friends." In *Ethics*, 107.2.  
 (2000) "Sidgwick's Minimal Metaethics." In *Utilitas* 12.3.
- SHOPE, R. (1978a) "The Conditional Fallacy in Contemporary Philosophy." In *Journal of Philosophy*, 75.  
 "Rawls, Brandt, and the Definition of Rational Desires." In *Canadian Journal of Philosophy*, 8.
- SIDGWICK, H. (1907) *The Methods of Ethics (Seventh Edition)*. London: Macmillan.  
 (1931) *History of Ethics, Sixth Edition (Enlarged)*. London: Macmillan.
- SKORUPSKI, J. (1989) *John Stuart Mill*. London: Routledge.  
 (1993) *English Language Philosophy, 1750-1945*. Oxford: Oxford University Press.  
 (1997) "Meaning, use, verification." In B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell Publishers.  
 (1999a) *Ethical Explorations*. Oxford: Oxford University Press.  
 (1999b) "Irrealist Cognitivism." In *Ratio*, 12.4.  
 (2000) "Desire and Will in Sidgwick and Green." In *Utilitas*, 12.3.  
 (2001) "Three Methods and a Dualism." In *Proceeding of the British Academy*, 109.  
 (2002) "The Ontology of Reasons." In *Topoi*, 21, 1-2.
- SOBEL, D. (1994) "Full Information Accounts of Well-Being." In *Ethics* 104.
- SOBER, E. (1992) "Hedonism and Butler's Stone." In *Ethics*, 103.1.
- STOCKER, M. (1976) "The Schizophrenia of Modern Ethical Theory." In *Journal of Philosophy*, 73.  
 (1990) *Plural and Conflicting Values*. Oxford: Oxford University Press.
- SUMNER, L. W. (1996) *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- VELLEMAN, D. "Brandt's Definition of 'Good'." In *Philosophical Review*, 97.
- WHITE, N. (1995) "Happiness in Aristotle's Ethics." In *Ethics*, 105.1.
- WHITE, S. A. (1990) "Is Aristotelian Happiness a Good Life or the Best Life?" In *Oxford Studies in Ancient Philosophy*, 8.
- WILKES, K. (1980) "The Good Man and the Good For Man." In A. Oksenberg Rorty (ed.), *Essays on Aristotle's Ethics*. Berkeley and Los Angeles: University of California Press.
- WILLIAMS, B. (1973) "A Critique of Utilitarianism." In J.J.C. Smart and B. Williams (eds.), *Utilitarianism and For and Against*. Cambridge: Cambridge University Press.  
 (1981) *Moral Luck*. Cambridge: Cambridge University Press.  
 (1985) *Ethics and the Limits of Philosophy*. London: Fontana Press/Collins.  
 (1993) *Shame and Necessity*. Berkeley & Los Angeles: University of California Press.
- WILLIAMS, B. & SEN, A. (eds.) (1982) *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
- WOOD, A. (1990) *Hegel's Ethical Thought*. Cambridge: Cambridge University Press.
- ZIMMERMAN, M. (2001) *The Nature of Intrinsic Value*. Maryland: Rowman and Littlefield Publishers.