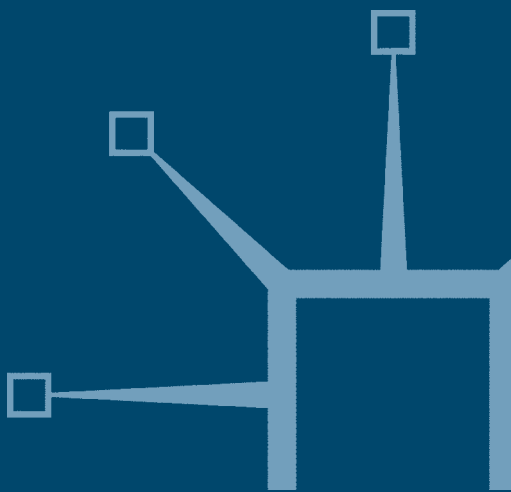
palgrave
macmillan

The Metaphysics of Autonomy

The Reconciliation of Ancient and Modern Ideals of the Person

Mark Coeckelbergh



The Metaphysics of Autonomy

By the same author

LIBERATION AND PASSION: Reconstructing the Passion Perspective on Human Being and Freedom

The Metaphysics of Autonomy

The Reconciliation of Ancient and Modern Ideals of the Person

Mark Coeckelbergh

palgrave
macmillan

First published 2004 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N. Y. 10010
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN 1–4039–3938–1 hardback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Coeckelbergh, Mark.
The metaphysics of autonomy: the reconciliation of ancient and modern ideals of the person / Mark Coeckelbergh.
p. cm.
Includes bibliographical references and index.
ISBN 1–4039–3938–1
1. Autonomy (Philosophy)–History. 2. Agent (Philosophy)–History. I. Title.
B105.A84C64 2004
128–dc22

2004046496

10 9 8 7 6 5 4 3 2 1
13 12 11 10 09 08 07 06 05 04

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham and Eastbourne

Contents

Preface

If I say that I want to be an autonomous person, what is it I want? Is it possible to construct a viable and acceptable ideal of the autonomous person, and if so, does this come at a price? This is the question I deal with in this book.

My argument is structured 'dialectically'. In Part I, I construct an ideal (thesis); in Part II, I discuss possible alternatives (antitheses), and that discussion results in my conclusion, my answer to the question (synthesis). I will also adopt this mode of presentation within Part I in my discussion of the contemporary and ancient ideals of autonomy. In Chapter 1, I articulate what I believe to be the modern ideal of autonomy (thesis); in Chapter 2, I present an apparent sharp contrast to this ideal, ancient ideals (antithesis). Then I develop a fresh ideal through engaging again with the thesis (Chapter 3) and with the antithesis (Chapter 4), trying to reconcile the modern and the ancient, thesis and antithesis.

My usage of the terms 'thesis', 'antithesis', and 'synthesis' to present the structure of Part I is not entirely orthodox. To the extent that ancient ideals of the person contrast with the modern ideal of autonomy, they are not competing views of the same thing. Strictly speaking, a view about what it is to be a person cannot be the antithesis of a view about autonomy. Furthermore, to the extent that the 'synthesis' does not incorporate key elements of both views, it is not a full synthesis. However, I will continue to use the terms since they are both views about what it is to be a person. Moreover, to the extent that 'thesis' and 'antithesis' can be reconciled, it is justified to speak of a 'synthesis'.

Let me further clarify my usage of terms. The term 'ideal' refers to something that is valued and aspired to. For example, an individual may value being a person or being autonomous. However, when I critically assess these values as ideals, I do not assess the values in themselves. When I use the term 'ideal' in phrases such as 'an ideal of the person' or 'the ideal of autonomy', I do not make a claim about whether being autonomous or being a person is valuable. Rather, I examine the *views* people have about what it is to be autonomous or being a person. Since I treat them as views, they can be, or cannot be, coherent views. The term 'an ideal of the person' refers to a view of what it is to be a person. It may also refer to a view of what a person

ought to be, or what it is to be a fully realised person; I will make this point in my introduction to the first chapter. The term 'the ideal of the autonomous person' refers to the view of what it is to be an autonomous person. Finally, with 'a dominant ideal' I mean that the ideal is accepted by most members of a society or culture, and that it is more influential than most other ideals accepted in that society or culture. For example, I will assume here that most people in what is commonly referred to as Western societies and cultures value personal autonomy, and I find it hard to think of another ideal that exercises such influence on the way people live and think in these cultures and parts of the world.

My choice of philosophical material to work with is, of course, precisely that: a choice. My sustained engagement with Harry Frankfurt – in Part I with his earlier work, in Part II with his later work – is an indispensable part of my book. Given that his ideas bear so strikingly on the issue I deal with, and that his thought has proved so helpful for my argument, I think this choice is fully justified. Furthermore, I have selected certain works of Plato and Augustine as key texts that represent 'ancient' (as opposed to modern) ways of understanding of what it is to be a person. Their particular relevance lies in the way they serve my purpose to create and resolve the (dialectical) tension mentioned above. I have chosen to use Sartre's idea of freedom for the same reason – at least what concerns the creation of tension. Finally, the presence of some of Kant's major works in my discussion not only is due to his being unavoidable in a book on 'the metaphysics of autonomy', but serves a specific role in the later part of my argument.

To conclude this preface, I wish to acknowledge my debts not mentioned elsewhere in the book. I am very grateful to Nicholas Dent and Iain Law. Their advice and suggestions have been of great assistance to me during the period of research and writing. Furthermore, I thank Robert Brecher and Mark Walker whose criticism on earlier work I have taken up as a challenge to write this book. Finally, I wish to thank in particular Christopher Warne and Paul Peterson who have commented on the final version of the manuscript. Their instructive comments have saved me from the most awkward phrases and have made me aware of some philosophical problems related to my usage of terms.

Part I

This page intentionally left blank

1
The Modern Ideal of Autonomy

1.1. Introduction

In contemporary society there are many different views of what an ideal person is. As I said in the preface, by 'ideal person' I mean the sort of person we wish to be, we aspire to. One of those ideals seems to be dominant: the ideal of a person which I roughly define as 'the autonomous person'. Most of us want to be, and to be seen as, beings that (are able to) govern themselves and their lives. We do not like other people telling us what to do, want, or be. We wish to (be able to) evaluate, decide, and do what we want. Moreover, this ideal is often linked to the belief that being autonomous is precisely being a person in the fullest sense of that word: if we attain autonomy, we realise ourselves as persons. The personal wish to have the capacity of, and to be in the state of, autonomy, is then also the wish to be a fully realised person. For example, if I wish to decide myself about which profession to take up, I may argue that as a person I have the capacity to make this decision myself, and that if I get the opportunity to use this capacity I get the chance to realise myself as a person in the sense of becoming more fully a person. Thus, the ideal of the autonomous person is a claim about what we are and ought to be.

The ideal of autonomy is present in contemporary philosophical discussions. Its presence *as an ideal* in the literature on autonomy is widely recognised. For example, in *Harm to Self*[1] (1986) Feinberg distinguishes between four meanings of autonomy,[2] autonomy as an ideal being one of them. However, the ideal also appears in discussions of related issues, and of freedom in particular. Whether it is expressed explicitly or not, discussions about freedom often illuminate the ideal of the autonomous person. Whether in discussions of freedom, free

will, or the nature of persons, we can identify a dominant normative position on what a person should be (i.e. an ideal of a person) that is either directly present in the discussion or indirectly motivates the arguments and the underlying assumptions. This can be illustrated by looking at the following influential contributions to recent philosophical discussions of freedom: Isaiah Berlin's distinction between two kinds of liberty, Christman's and Feinberg's definitions of autonomy, Harry Frankfurt's hierarchical model of the volitional structure of a person, and Charles Taylor's concept of 'strong evaluation'. I will use the ideas of these contemporary authors to outline the modern ideal of the autonomous person.

My answer to the question 'What do we mean if we say that we aspire to be autonomous persons?' will constitute the first step in the dialectical structure of my argument in Part I: the thesis.

1.2. Berlin, Christman, and Feinberg

In his famous essay 'Two Concepts of Liberty' (1958), Isaiah Berlin distinguishes two senses of freedom or liberty (he uses both words interchangeably). He first characterises a 'negative' sense as an 'answer to the question "What is the area within which the subject – a person or group of persons – is or should be left to do or be what he is able to do or be, without interference by other persons?"'. Secondly, he develops a 'positive' sense as an 'answer to the question "What, or who, is the source of control or interference that can determine someone to do, or be, this rather than that?"' (Berlin 1997 (1958): 194). Of the two senses, the 'positive' sense of freedom is particularly revealing with regard to the content of the contemporary ideal of autonomy:

> The 'positive' sense of the word 'liberty' derives from the wish on the part of the individual to be his own master. I wish my life and decisions to depend on myself, not on external forces of whatever kind. I wish to be the instrument of my own, not of other men's, acts of will. I wish to be a subject, not an object; to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from outside. I wish to be somebody, not nobody; a doer – deciding, not being decided for, self-directed and not acted upon by external nature or by other men as if I were a thing, or an animal, or a slave incapable of playing a human role, that is, of conceiving goals and policies of my own and realising them. This is at least part of what I mean when I say that I am

> rational, and that it is my reason that distinguishes me as a human being from the rest of the world. I wish, above all, to be conscious of myself as a thinking, willing, active being, bearing responsibility for my choices and able to explain them by reference to my own ideas and purposes. I feel free to the degree that I believe this to be true, and enslaved to the degree that I am made to realise that it is not. (Berlin 1997 (1958): 203)

One of the first things we notice about this passage is Berlin's repeated reference to the *wish* on the part of the individual to be his[3] own master, a subject, a willing, active being. This reference to wishes indicates that to be this kind of individual is thought to be an *ideal*, something the individual wishes and aspires to be(come). Secondly, we can see that the ideal Berlin describes is an ideal of *autonomy* because it captures what Christman calls the 'core' meaning of autonomy. This is the meaning Feinberg's distinct conceptions centre on: 'the actual condition of autonomy defined as a psychological ability to be self-governing' (Christman 1989: 5). In other words, Berlin's conception of 'positive' freedom presents us with an excellent picture of the ideal of a person as an autonomous individual; and it is precisely this ideal that is dominant in the literature. Berlin's description of 'positive' freedom suggests that discussions about autonomy focus on questions about the source of control and determination. The ideal of the autonomous individual suggests the following answer: the source should be *me*, and not something or somebody else. *I* should determine myself; not something or somebody else. In other words, I exemplify this ideal of what a person should be when I am this autonomous individual, when the source of my action and my thought is me, when I determine myself as an individual. What we mean when we call this an 'ideal' is simply that it is a state or condition to which an individual should aspire. In his essay 'The Idea of a Free Man' Feinberg puts it like this: 'I am autonomous if I rule me, and no one else rules I' (Feinberg 1973: 161). This is the core content of the ideal of a person present in Berlin's more elaborated description.

1.3. Frankfurt

If we say that we want to be self-determining, part of what we mean is that we (want to) see ourselves as different from 'the world' with its deterministic character; if we understand the world as fully determined by the laws of science, we want to say that we're different

from that world since we have a free will. In 'Freedom of the Will and the Concept of a Person' (1971) Frankfurt identifies one essential difference between persons and other creatures in the structure of the person's will: persons are able to form second-order desires. 'Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are' (Frankfurt 1982 (1971): 82). Frankfurt then makes a distinction between two different kinds of agents: persons and wantons. Frankfurt uses the term 'wanton' to refer to an agent who does not care about his will. This means that 'his desires move him to do certain things, without its being true of him either that he wants to be moved by those desires or that he prefers to be moved by other desires' (Frankfurt 1982 (1971): 86). The key difference between persons and wantons is that although wantons may have second-order desires (to want to have a certain desire), only persons are able to form what Frankfurt calls second-order *volitions*: the person 'wants a certain desire to be his will' and, as I interpret this requirement, succeeds in having that will (Frankfurt 1982 (1971): 86). A person is able to do this only in virtue of his rational capacities. 'For it is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will and of forming volitions of the second order. The structure of a person's will presupposes, accordingly, that he is a rational being' (Frankfurt 1982 (1971): 87).

Whether Frankfurt's distinction between second-order desires and second-order volitions (and therefore between wantons and persons) is convincing or not,[4] it is clear that his discussion embodies a certain ideal of a person. For many of us, Frankfurt's account of a person is not only a description of what persons essentially *are* (that is, we agree with his view about what persons are); it is also an ideal by which we wish to live as persons. We want to be able to decide whether we endorse a first-order desire or not, we want to be able to choose between various (first-order) desires. We want to 'want a certain desire to be our will' and succeed in having that will (Frankfurt 1982 (1971): 86). We want to succeed, not just want; we want to be in control. Otherwise, to use Frankfurt's terminology, we are a wanton, not a person, and this we do not wish to be. We place a high *value* on persons in Frankfurt's sense – we value the idea of a self-determining individual able to form higher-order desires and govern his first-order desires by them.

1.4. Taylor

Charles Taylor goes further than Frankfurt in arguing that persons not only have the capacity to question their (first-order) desires, but also the capacity to question themselves. In 'Responsibility for Self' (1976) he argues that

> if we think of what we are as defined by our goals, by what we desire to encompass or maintain, then a person on this view is one who can raise the question: Do I really want to be what I now am? (i.e. have the desires and goals I now have?) In other words, beyond the *de facto* characterisation of the subject by his goals, desires, and purposes, a person is a subject who can pose the *de jure* question: is this the kind of being I ought to be, or really want to be? (Taylor 1976: 111)

This passage specifies Taylor's conception of the ideal of a person: that he should ask himself whether this is the kind of individual he ought to be, or really wants to be. We not only have the capacity to ask these questions; many believe (exercising) this capacity is essential to what we are. Taylor claims that 'we have the notion that human subjects are capable of evaluating what they are' and that 'many' believe this kind of evaluation to be 'essential to our notion of the self' (Taylor 1976: 112). According to Taylor, we as a matter of fact make judgements and 'strong' evaluations. We do not just evaluate what to do in the sense that we desire to do this rather than that, but we also evaluate whether it's good or not to have this desire. Taylor distinguishes between 'someone who evaluates non-qualitatively, that is, makes decisions like that of eating now or later, taking a holiday in the north or in the south' and someone who 'deploys a language of evaluative contrasts ranging over desires'. Taylor refers to the latter type of a person as a 'strong evaluator' (Taylor 1976: 116). He argues that the capacity for strong evaluation is 'an essential characteristic of a person', since beings other than persons (such as animals) are either incapable of evaluating desires or are only able to evaluate as 'a simple weigher' (Taylor 1976: 117–18).

Taylor's idea of strong evaluation is compatible with Frankfurt's idea of higher-order desires. In particular, the capacity to question whether I am now the person I really want to be can manifest itself in the formation of higher-order desires, such as the desire to be a different person. In Frankfurt's volitional account this amounts to the desire to have a different volitional structure.

Taylor explicitly refers to something like 'ideals of a person'. He suggests that such ideals are operative in making 'strong' evaluations: we aspire to be a certain kind of person, and this influences our evaluations, our judgements. In particular, Taylor describes the following ideal: we should be a concerned person, concerned about the issues touching the quality of our lives which seem to us basic or important (Taylor 1976: 117). In the light of the question of autonomy, this means that we should be self-determining in the second aspect I distinguished in my introduction: we should be masters of our lives. The suggestion is that I should evaluate myself on the level of my life as a whole; I should question the way I live my life, the choices I make, the things I do. Taylor's account, therefore, adds to what we already know from Frankfurt's. To exercise my capacity to determine my life means that I make 'strong' evaluations. Strong, because I do not just classify a certain (first-order) desire as good or bad, but I question myself as a certain kind of person. Taylor calls this questioning of the self 'radical evaluation', a 'reflection about the self' which 'engages the self most wholly and deeply'; at stake here is my identity (Taylor 1976: 126).

1.5. Further refinements

In the beginning of this chapter I claimed that autonomy is nowadays a very influential ideal in the sense that it is something many of us aspire to, and that many believe autonomy is a capacity of persons (alone) which needs to be exercised if we want to be fully realised persons. It is not my purpose here to establish the empirical validity of this claim; others are welcome to challenge it if they wish. Even if it were to turn out less important than I assume here, most readers will agree with me that (the ideal of) autonomy is at least of sufficient importance and influence in contemporary Western society and culture to merit and motivate extensive philosophical analysis and reflection.

What is autonomy? So far I have used 'autonomy' to refer to the capacity and exercise of self-control and self-government, of deciding yourself about your desires and your life. I have said that this involves questioning and evaluating your desires, yourself, your identity. But much of what has been said raises questions. Firstly, we could consider some broad questions about autonomy. What does it mean to say that a person decides about his life? Is the ideal of autonomy meaningful if we have little or no options available to us? Is autonomy a capacity or an achievement? What about autonomy as a political concept? Does

self-government mean that I can do what I want? How does an autonomous person relate to other (autonomous) persons? And what is the relation between autonomy and morality? Secondly, there are more specific questions about my interpretation of Frankfurt and Taylor. For example, the reader may wonder what it means 'to want a certain desire to be your will' or 'to question your identity'. Many of the latter sort of questions will receive further clarification in the course of the book, in particular in the sections on Frankfurt. The first broader questions I will consider now. I will not be able to fully answer them here, but they will help me to refine my construction of the modern ideal of autonomy. Consider the following distinctions and points of clarification.

1.5.1. My desires and my life

I would first like to make a distinction between two senses of 'self-government', which, recall, Christman identifies as the core meaning of autonomy. To start mapping the area where the notion of self-government is applicable, we can make a distinction between, on the one hand, the issue of who is in control or who determines my desires, and, on the other hand, the issue of who is in control of or who determines my life as a whole. Discussions of freedom typically relate to both aspects. For example, Frankfurt claims that freedom requires, as we have seen earlier, the ability to form second-order desires: 'Besides wanting and choosing and being moved to do this or that, men may also want to have (or not to have) certain desires and motives. They are capable of wanting to be different, in their preferences and purposes, from what they are' (Frankfurt 1971: 82–3). This can be understood as the exercise of control over a first-order desire, or as making a choice between two conflicting first-order desires. But the result of such exercise(s) may extend beyond the present moment. If I change my preferences and purposes, my actions will probably change as well, and therefore the course of my life may change. Thus, the above distinction between control over my desires and control over my life concerns the range of the effect of the exercise of control or choice. The effect may be limited in time, for example, to this very moment when I decide to control my desire to eat, or may extend to my life as a whole, when the interplay between my desires of a higher order and my first-order desire(s) results in an effective change in the course of my life. It gives us a more complete picture of the content of the ideal to distinguish between the ideal of a person controlling his present

desires, on the one hand, and the ideal of a person directing his life. This second aspect of the ideal of a person dominant in our society and culture is at least as much present in the literature as are discussions of self-control in the more limited sense just defined. It is held, for example, that for a person to direct his life it is good (ideal) to have plenty of options available.

1.5.2. The ideal of having many alternatives

The ideal of having plenty of options available is different from the ideal of autonomy, since it is more concerned with the setting or context in which the person is choosing and acting. However, it shows concern with self-direction as the direction of one's life as a whole. In 'On the Necessity of Ideals', Frankfurt writes that 'our culture places a very high value on a certain ideal of freedom according to which a person is to have varied alternatives available in the design and conduct of his life' (Frankfurt 1999: 108). Firstly, although he writes about an ideal of *freedom*, it seems clear that this ideal of freedom is closely linked to the ideal of a person: one who has many options. It seems that when the ideal of having many options available is realised, the pursuit of the ideal to design and conduct one's own life becomes more meaningful. After all, if we have but one option available, we arguably have no real 'option' at all, and there is nothing left to design or to conduct for oneself. Therefore, the ideal of designing one's own life presupposes the availability of alternatives, and, ideally, many of them. Secondly, Frankfurt is suggesting that this ideal of freedom – and therefore also this ideal of a person – is indeed an *ideal*, or at least a very influential idea in society: our culture places 'a very high value' on it.

1.5.3. Capacity and condition; political autonomy

Furthermore, using Feinberg's distinction between the *capacity* for self-government and the actual state or *condition* of self-government,[5] we could make a distinction between the ideal of having the capacity to govern oneself, on the one hand, and the ideal of being in a state or condition of self-government, on the other hand. A person who aspires to be autonomous might refer to the capacity, the condition, or both. Finally, it is also possible that the person aspiring to be autonomous refers to what we may call 'political' autonomy, which means the government of one's own 'domain' (Feinberg 1986: 28). I take this to refer to self-government as a person, being able to choose and act within a certain sphere, comparable with what it means for a state to

be autonomous. Although the latter meaning will not be completely excluded (as Feinberg observes: the different meanings of autonomy *are* closely related), the primary object of this study is the ideal of autonomy in the first two meanings, the ideal of having the capacity to govern oneself, and the ideal of being in the condition of actually governing oneself. My discussion will make further refinement necessary, but at present I shall hold on to this distinction.

1.5.4. The ideal of 'doing what you want'

If I have been speaking of *the* dominant modern ideal, this may misleadingly suggest that this ideal is so overwhelmingly dominant that there is little room for other modern ideals. It seems to me that there is at least one other ideal which is perhaps not dominant but deserves our consideration since it could be considered as a 'rival' ideal. In addition to the dominant modern ideal of a person as an autonomous individual I have articulated so far, there is the modern ideal of an individual who has the capacity to do, and is in the condition of doing, whatever he wants to do. I shall refer to the latter ideal as 'the ideal of doing what you want', 'doing what you want' meaning unlimited freedom of action. This ideal is arguably not less influential in modern society than the former, has been present and popular in premodern times too, and has been subject to criticism throughout the history of human thought. I shall maintain that there is a clear distinction between the dominant ideal of a person as someone choosing and acting as an autonomous individual and the ideal of 'doing what you want'. There are two main sources to a denial of this distinction. Firstly, it is common in contemporary society to uncritically ignore the distinction and simply equate personal autonomy with 'doing what you want'.[6] Against the view that autonomy is unlimited freedom of action, we must clarify philosophically the meaning of autonomy and show that and why it is a distinct ideal from 'doing what you want'. Secondly, apart from the ideas around in contemporary society, there is a philosophical strand and tradition of thinking about freedom and about autonomy in terms of 'doing what you want' (Hobbes, Hume, Mill, etc.[7]). In opposition to these views I will affirm the modern ideal of autonomy as an ideal distinct from 'doing what you want', for it is as such that the modern ideal of autonomy is a significant modern ideal. It is a significant modern ideal in the sense that it is (1) a well-defined ideal, (2) a dominant ideal today, and (3) an ideal that can be defended adequately as a cogent philosophical notion. This focus on the most significant ideal of autonomy does not mean that from now

on the issue of 'doing what you want' is excluded from the discussion altogether. It will remain 'in play', since often it helps to say what the ideal of autonomy is *not*, and it is therefore of assistance in the process of clarification by contrast. So although the ideal of 'doing what you want' has not been clearly defined yet, I will do this in further discussion,[8] which will bear out the precise philosophical issues at stake here, the precise points where the difference(s) between the two ideals lie(s).

It could be argued that it is not necessary to speak of two different ideals here, borrowing Feinberg's distinction between concepts and conceptions[9] (Feinberg 1986: 27–8). Feinberg distinguishes between different conceptions of the concept of autonomy, by which he means different (inter-related) meanings of this concept. The ideal of autonomy is one such conception – a conception of the concept of autonomy. In turn, this conception can be regarded as a concept, having itself different possible conceptions (meanings, interpretations) of it. Thus, if the concept I am dealing with is the ideal of autonomy, then 'doing what you want' is a conception of this ideal. However, calling it a conception may suggest that it is a possible, viable, and valid interpretation of autonomy, a legitimate child of the mother concept. Therefore, if we do not want to speak of distinct ideals, it would be better to apply to 'doing what you want' either a neutral term (interpretation) or a term relatively biased in favour of the ideal of autonomy as against 'doing what you want'. I affirm both the distinction of autonomy from 'doing what you want', and the primacy of the former as an ideal over the latter. Since I argue for these positions throughout this and the following chapters, I will use the seemingly 'biased' way of characterising their relationship. In what follows, the starting point is the argument *that* there is a distinction, which will then initiate a discussion of how both distinctive elements are precisely related.

The distinction between the modern ideal of the person as an autonomous individual and what I will now call the 'poor' variant or wrong interpretation of this ideal, i.e. the ideal of doing what you want, can be clarified by looking at contemporary discussions of freedom. I think Frankfurt's distinction between freedom of action and freedom of will is particularly helpful here:

> A person who is free to do what he wants to do may yet not be in a position to have the will he wants. Suppose, however, that he enjoys both freedom of action and freedom of the will. Then he is not only free to do what he wants to do; he is also free to want what he wants to want. It seems to me that he has, in that case, all

> the freedom it is possible to desire or to conceive. There are other good things in life, and he may not possess some of them. But there is nothing in the way of freedom that he lacks. (Frankfurt 1982 (1971): 93)

So Frankfurt makes a distinction between freedom of action and freedom of will. Freedom of action means being able to do what you want. But the quotation above suggests that this is not 'all the freedom it is possible to desire or to conceive' and so it cannot be an absolute or unqualified ideal. For, Frankfurt's second kind of freedom, the freedom of the will, allows for a kind of 'overriding' mode of freedom, of being able to regulate 'doing what you want', by not wanting a particular first-order desire: I may desire to not want something I want. For example, the drug addict may want to take his drug, but at the same time may not want to want the drug, may not want to have this desire. He has the second-order desire to not have the first-order desire. Although Frankfurt (at least in his earlier work) does not explicitly mention any *ideal* (his discussion is a discussion of freedom), he argues that I am only fully a person if I am able to develop and exercise this second-order desire to want or not want a certain desire to be my will (see earlier distinction from animals and wantons (Section 1.3.). It is *this* freedom, the 'freedom of will', and not (only) the 'freedom of action' which is held to be an ideal, and which is arguably (part of) the ideal of the autonomous person.

This is the distinction explained by using Frankfurt's conceptual apparatus; there are other ways to explain the ideal of autonomy in contrast with the 'ideal' of 'doing what you want', and I will refer to other writers and currents of thought later in this book. As I have suggested already, although it is not my main purpose to defend the ideal of autonomy *against* the 'ideal' of 'doing what you want', it is necessarily part of this book. As will be shown, the view that 'doing what you want' is an ideal of the person suffers from the following problems.

Firstly, it is internally inconsistent, so not an ideal. The pursuit of 'doing what you want' as an ideal inevitably undermines itself, since empirically any effort to devote yourself exclusively to 'doing what you want' puts you in a position in which you cannot do what you want. To use a metaphor: you become the slave of your desires. The pursuit of the ideal of 'doing what you want' is ultimately self-defeating. Apart from the empirical argument, we can also argue for this point by using our understanding of Frankfurt's ideal of a person (and, as we will see, the extended version of this ideal that will be developed in this book).

The gist of it is the idea that to really be autonomous you need to direct yourself to something which is not simply the objective of your own first-order desires. Only then it is possible to 'identify decisively' with a certain desire, to make it 'your own'. Otherwise there is loss of control and a certain desire will 'enslave' you. (I will present this argument in due course (Chapters 3 and 4)).

Secondly, 'doing what you want' is arguably not a dominant ideal. Who seriously holds and practises this ideal – with its full consequences? To practice 'doing what you want' would imply that we always follow our desires, which is inconsistent with what we actually do. We know that we are not everywhere and always *able* to do what we want. For example, the mere fact that we live in society restricts our freedom to do what we want. It could be objected that although we are not able to realise the ideal, we could still wish to do what we want, it could be still our ideal. However, this is inconsistent with our beliefs: most of us believe that we also *ought* not always to follow our desires.

Thirdly, 'doing what you want' is a misinterpretation of, and so is distinct from, the ideal of autonomy (I began this argument earlier in this chapter). The grounds for the validity of this claim will be demonstrated parallel with the work done on the refinement of the ideal of autonomy; the clearer this ideal becomes, the more obvious its difference from 'doing what you want' will be. As noted above (see note 8 to this section), I will also explicitly point out and discuss this difference when appropriate.

1.5.5. Inner and outer autonomy

Feinberg seems to agree with the point that 'doing what you want' does not deserve the name 'autonomy'. In his essay 'The Idea of a Free Man' (1973) we find autonomy defined as 'I am autonomous if I rule me, and no one else rules I'. Although I have quoted this definition before, closer inspection reveals two distinctive aspects of autonomy, which I shall call an 'outer' aspect and an 'inner' aspect. This distinction is very important and will be a significant feature of the forthcoming discussion.

What I mean by the inner aspect of autonomy is captured in the phrase 'I rule me'. The ideal of autonomy here is that we are able, in Feinberg's words, to 'identify with the desire that is higher in our personal hierarchy, and consider ourselves to be the subject rather than the object of constraint' (Feinberg 1973: 148). This ideal can be refined by looking at what it would be for a person *not* to be autonomous. According to Feinberg, the non-autonomous person is a person who

has 'no hierarchical structure of wants, and aims, and ideals, and no clear conception of where it is within him, that he really resides'. Such a person would be 'a battlefield for all of his constituent elements, tugged this way and that, and fragmented hopelessly' (Feinberg 1973: 149). Although such a person may have authentic desires and aims, desires and aims that can be called his *own*, he fails to possess and exercise autonomy because the internal order and structure is lacking. Feinberg argues that even though such a person may 'do anything he wants', he is unable to order the options available in a hierarchy. Feinberg concludes that 'surely it is more plausible to construe such a state as unfreedom than as an illustration of the dreadfulness of too much freedom' (Feinberg 1973: 149). This shows how ambiguous the notion of freedom is. Autonomy, by contrast, lends itself to a more precise definition, and is therefore a better candidate to use as the key concept of the modern ideal of a person. It is clear that this person – as characterised by Feinberg – is not autonomous, whatever his status in terms of freedom. 'When the *I* is incapable of governing its *Me*, the result is anomie, a condition which is not control from without, but rather being virtually "out of control" altogether' (Feinberg 1973: 163).

Whereas the inner aspect of autonomy refers to the relations and state of affairs *within* the person, I shall say that outer autonomy concerns the relationship between the person and the rest of the world, in particular other people. Indeed, I hold that since the laws of the natural world are beyond our control (we can only try to discover them and use that knowledge), the issue of autonomy in its external aspect arises only in relation to other people, and the social and cultural world in general. Again it is useful here to define autonomy by looking at one of its contrasts: 'forms of passive mindless adjustment (the pejorative term is "conformity") to the requirements of one's culture' (Feinberg 1973: 163). Here the other-direction is total; there is complete attunement to the wishes of others. Autonomous persons, on the contrary, are 'capable of conforming if they choose. […] They will conform when and only when there are good reasons for doing so; and they can attend to reason free from the interfering static of "signals" from other voices. They can control the speed and direction of their gyroscopes… [They are not] indifferent to the reactions of others, but [they] *can* be moved by other considerations too' (Feinberg 1973: 165).

What precisely is the relationship between the inner and outer aspects of autonomy? To arrive at a complete picture containing both aspects, we can use Feinberg's comparison between the independence and self-government of the individual and the independence and

self-government of the state. He remarks that 'self-government might turn out to be more repressive even than foreign occupation. Yet, for all of that, the state might still be politically independent, sovereign, and governed from within, hence free. Analogously, it is often said that the individual person is "free" when his or her ruling part or "real self" governs, and is subject to no foreign power, either external or internal, to whose authority it has not consented' (Feinberg 1973: 158–9). So here we have two aspects together, the external and internal one, the outer and inner one. Only the outer is not enough for autonomy to be complete; there could be some form of 'inner' repression which prevents the person from attaining full autonomy. (I shall discuss 'inner autonomy' in the course of this book; I will return to the issue of repression in particular in Part II, Section 6.4.1.)

Feinberg reminds us that 'free' means also having a certain legal-political status. Historically, to be a freeman was to be a full member of the political community. One contemporary meaning of political freedom as being entitled to certain rights on the basis of membership of the state as a political community still relates to that historical meaning. However, we may say that the ideal of the autonomous person as outlined before does not depend on this political freedom. It could be imagined that somebody is literally a slave but is still able to order his desires and aims in a hierarchy. Although he is not able to do or get what he wants, he may be an ideal person in the sense that he has the balanced order of autonomy, residing completely in himself. However, it may be objected that the modern ideal of the autonomous person does include 'not being a slave', since this would mean that he lacks 'outer autonomy'. Therefore, I conclude that autonomy is not a necessary condition for political freedom, but political freedom is a necessary condition for autonomy. However, I argue that although *full* autonomy is not possible in the condition of slavery, a slave *might* still possess inner autonomy. If I can choose whether to do what I have to do *willingly* or not, making the command my *own* wish or desire or not, I am still autonomous in the 'inner' sense. So although I'm not free and not autonomous in the 'outer' sense I do not lack 'inner' autonomy. To that extent, it may be still possible for me to realise part of the ideal of a person, whatever my legal or political status is.

This discussion shows that the distinction between 'inner' and 'outer' autonomy allows us to further analyse the meaning of autonomy and to make other distinctions which remain concealed if we applied the term 'autonomy' on its own. Equally, the term 'freedom' on its own appears to be too vague to allow for such distinctions. We

have now a clear criterion to decide whether a person is autonomous in the fullest sense of the word, namely two necessary conditions. The first condition is that I am not restrained by, and not dependent on, something inside me. The second condition is that I am not restrained by, and not dependent on, something outside me. Whether a certain condition or state of the person satisfies these criteria depends on (a) whether or not something can be identified as putting a constraint on me ('Is it really a constraint?' and 'Am I really dependent on it?' are the questions to be asked); and on (b) whether or not this constraint or source of dependence – *if* there is one at all (question a) – is something inside or outside me ('Is it really an 'inner' constraint?' and 'Is it really an 'outer' constraint?' are the key questions).

This criterion, based on the distinction between 'inner' and 'outer' autonomy, will be used throughout this book as an analytical tool. As an aspect of the argument about autonomy, it will be consolidated through this use, and its explanatory power will be shown more fully as the general argument of my book develops.

1.5.6. Autonomy and morality

Does autonomy include the freedom to choose evil? Considering the dominant ideal of autonomy as articulated so far, it is evident that the authors discussed share the view that autonomy and morality are to be considered as fundamentally distinct issues. Whether or not I choose evil, if it is *my* choice, a choice I take in the capacity of being my own master, a subject, a doer, a willing being (Berlin), if I am self-governing (Christman) and self-ruling (Feinberg), if I succeed in having the will I want (Frankfurt), if I engage in strong evaluation (Taylor), then, on this view, there is nothing in the way of autonomy that I lack. Feinberg explicitly draws a distinction between the ideal of autonomy and moral excellence, arguing that since autonomy is consistent with ruthlessness, cruelty, and other 'failings', it is at best only a partial ideal 'insufficient for full moral excellence'(Feinberg 1986: 45).

It remains to be seen whether this view can be defended as coherent. I will return to this issue in Chapter 4, where it will receive further discussion in the light of my attempt to reconcile 'modern' and 'ancient', and in Part II (for example, in my discussion there of Frankfurt). But it seems to me that there is no doubt that the view that autonomy, by itself, is morally 'neutral', and that an autonomous person can choose evil, is a key part of the *dominant* modern ideal of autonomy. There may be other views, but this is the dominant one and the one I choose to engage with in this book.

1.6. Conclusion: A sketch of the modern ideal of the autonomous person

My discussion of autonomy so far yields a reasonably clear picture of what it means to say 'I wish to be an autonomous person'. It means that I wish to rule, govern, and determine myself. This ideal has an 'inner' aspect (I wish to rule my desires) and an 'outer' aspect (I wish not to be ruled by something outside me). If I want the latter, I also want political liberty. Furthermore, the ideal includes the wish to exercise my *capacity* to rule myself or the wish to be in the *condition* of self-rule. It can concern my present (inner or outer) situation or extend to the future and my life as a whole. It can include the wish to (be able to and be in a condition to) use my governance and rulership to evaluate and question my desires or perhaps even myself as a person. In Frankfurt's and Taylor's words, I wish to form higher-order desires (volitions) or engage in strong evaluation. Related to this wish is the ideal of being a *person*, in the sense of wishing to be different from animals or wantons, who lack this capacity of self-evaluation (inner aspect), and in the sense of wishing to have and to use the capacity of free will, to be a space of freedom in a deterministic universe (outer aspect). This ideal is different from the ideal of doing what you want; if I rule myself I constrain myself in some way. But this constraint is not necessarily moral. The dominant modern ideal of autonomy is construed as a morally neutral ideal, that is, the wish to be autonomous does not necessarily include the wish to be morally good or excellent; the wish to be autonomous and the wish to do evil are seen as consistent.

Although not all of us share this ideal of the autonomous person, there is no doubt that it occupies an influential place in our contemporary culture and society. This was not always the case. In the following chapter I will consider 'ancient' (as opposed to 'modern') ideals of the person that appear to be very different from, if not entirely opposed to, the modern ideal of autonomy.

2
Ancient Ideals of the Person: Plato and Augustine

2.1. Introduction

I have completed now the first step in the development of an ideal of autonomy by articulating what I believe to be the dominant modern ideal of autonomy. In terms of the dialectical structure of Part I, this means that the *thesis* is constructed. In this chapter I turn my attention to the construction of an *antithesis*. As the title suggests, I have chosen 'ancient ideals' as a name for certain elements in Plato's and Augustine's thought that contrast sharply with the modern ideal of autonomy. Why have I chosen to concentrate on Plato and Augustine? How do elements in their thought contrast with the modern ideal, and how sharp is this contrast really? And why is this contrast relevant to the question concerning the ideal of *autonomy* anyway? Most of these questions cannot be fully answered in this chapter alone. For example, the full relevance of Plato and Augustine for the modern ideal of autonomy will only become clear in the course of the following chapters; and the critical question concerning the sharpness of the contrast will be mainly dealt with in Chapter 4. But I hope here to construct a convincing antithesis that serves the further development of my argument, that is, builds up the tension necessary for the later construction of a synthesis and avoids resolving the tension too soon by anticipating the synthesis.

I will start with a brief explanation of my choice of Plato and Augustine, and a short exposition of the contrast elements in their views provide with the modern ideal of autonomy. Then I will analyse these elements in Plato and Augustine to reconstruct ancient ideals of the person as the antithesis of the modern ideal of autonomy.

The question central to this book stems only partly from concerns arising out of contemporary discussions of freedom and autonomy.

A second source of the issue I want to deal with is the observation that there are in the history of philosophy descriptions of phenomena and theoretical reflections that embody or are motivated by what at first sight seem to be very different ideals of a person. Although the Platonic and Augustinian ways of understanding freedom and ideals of the person are part of the Greek and Christian roots of modern civilisation, their ideas seem in many ways hostile to the ideal of the autonomous person we cherish today. Consider the following two text fragments, one from Plato and one from Augustine:

FRAGMENT ONE

'The greatest of goods come to us through madness, provided that it is bestowed by divine gift.' (Plato, *Phaedrus* 244a/57)

FRAGMENT TWO

'But you will free me, O Lord; I know that you will free me. For ever I keep your mercies in mind.' (Augustine, *Confessions* X.34/241)

These two fragments suggest a completely different form of self-understanding that contrasts with the dominant modern way of thinking about a person that I have outlined in Chapter 1. Without further reflection, and on the basis of these fragments alone, it seems to be impossible to construct a description – let alone an ideal – of a person that makes sense to *us*, living with the dominant ideal of the autonomous individual. Of course we would need more information, more primary and secondary sources. But the main problem here is of a different kind: we do not generally think of a person in these terms. Although today many people are religious, modern culture and philosophy generally tend to discourage thinking about ourselves and the world in a divine context, and many of us simply can't imagine how it is (a) to be mad and receive a divine gift, or (b) to become free by the mercies of God, let alone that either of those would be an *ideal* of the person. We wish to be in control, we want to determine ourselves and our lives; in other words, we hold the ideal of the person as an autonomous individual. The ideas that the greatest goods are given to us by gods or that it is up to the mercy of God whether we are free or not appear to be in sharp contradiction to this ideal. It seems that there is an unbridgeable gap between the ideals – if the ancient descriptions represent any 'ideals' at all.

To bring this out more fully, I will now articulate more fully the Platonic and the Augustinian ideals of the person and show why these

ideals are attractive and interesting but nevertheless appear to be irreconcilable with the modern ideal. For that purpose I select Plato's *Phaedrus*, Book VII of the *Republic*, Augustine's *On Free Choice of the Will*, and the *Confessions*.

2.2. Plato's ideal of the person

2.2.1. The *Phaedrus*

To articulate Plato's ideal of the person in the *Phaedrus*, I will focus on (1) his metaphor of the charioteer and the horses, (2) his view of madness, and (3) discussions of Dionysian madness and madness in ancient Greek culture, the better to understand and further refine the ideal constructed using the former two elements.

The charioteer and the horses

In the *Phaedrus* Plato compares the soul to a chariot being pulled by two horses, with the charioteer controlling the horses or failing to do so (246a–257a). One horse represents appetite. The second horse represents restraint, and – as we will see later – is related to (what Plato means by) love. The charioteer (reason) tries to manage the horses in the light of what vision he has of the good[10] (see also Dilman 1999: 36). Plato says the following of the horses:

> The right-hand horse is upright and cleanly made; he has a lofty neck and an aquiline nose; his colour is white, and his eyes dark; he is one who loves honour with modesty and temperance, and the follower of true opinion; he needs no touch of the whip, but is guided by word and admonition only. The other is a crooked, lumbering animal, put together anyhow; he has a short, thick neck; he is flat-faced and of a dark colour, with grey eyes and blood-red complexion; the mate of insolence and pride, shag-eared and deaf, hardly yielding to whip and spur. (253d–e)

Now from the way Plato paints this picture we can infer his ideal of the person: the ideal person is the one who exercises self-control, restraint, modesty, etc. and is able to do this because he has reason and exercises it. So far, this picture is roughly compatible with the modern ideal. Using Frankfurt's terminology, we can interpret this ideal as the wish to form second-order desires. I do not want to be 'enslaved' by my first-order desires but exercise my capacity of self-rule. According to the modern ideal of autonomy, however, this does not necessarily mean

that I always restrain the desires Plato calls 'appetite'. Rather, it means that I wish to decide myself whether I identify with those desires or not. I want to be in control. Thus, there seems to be a difference from Plato, since he seems to prefer restraint over appetite. It appears that in Plato's view we have to use our capacity of self-control to restrain our appetites. Certainly, the modern ideal is often combined with this preference for self-restraint, but I want to make a distinction between the 'neutral' capacity and condition of self-rule that is always part of the modern ideal of the autonomous person, on the one hand, and the particular preference for self-restraint and modesty that sometimes accompanies it.

Furthermore, the ideal of self-mastery contained in Plato's metaphor can be interpreted as an ideal of 'inner autonomy'. Dilman reminds us that self-mastery should not be understood here as 'using the whip' on the horse representing appetites. Rather, the soul is transformed into a team that acts in harmony. This, Dilman says, is the correct meaning of autonomy: 'It is in such self-mastery that the person will have achieved *autonomy*: what he does will be what he wants to do, not what he is forced to do, and he will be wholly behind it' (Dilman 1999: 39).

However, there seems to be a more fundamental difference between Plato's ideal and the modern ideal. There is a very important part of the picture which should not be overlooked: the charioteer (reason) is able to manage the horses *if and only if he has a vision of the good*. It is important to see that this inner autonomy as harmony or wholeness comes about only as a result of submitting to, and being moved by the good. Such a commitment, namely a commitment to a metaphysical entity called 'the good' is unacceptable to most people holding the modern ideal of autonomy. Many of us understand autonomy as involving independence from such things as 'the good'. Consider the following argument. To be able to rule myself I do not need 'the good'. Moreover, if I were dependent on 'the good' for my autonomy this would make it impossible to rule myself, since then I would be ruled by 'the good' rather than myself. Therefore, a commitment to something like 'the good' is neither necessary nor sufficient for my autonomy. I will challenge the validity of this argument in the next chapters, but for now it appears to make sense given my articulation of the modern ideal. Recall that in section 1.5.5. I have defined not being dependent on something outside me as a necessary condition for 'outer autonomy'. It looks as though the modern ideal of autonomy is (largely) incompatible with Plato's ideal of the person. Any attempt at reconciliation would have to overcome this problem.

Plato and madness

There is, however, another ideal present in Plato's writings that is, in my view, much more alien to the modern ideal. In my introduction, I quoted his claim about the value of madness: 'The greatest of goods comes to us through madness, provided that it is bestowed by divine gift' (Plato, *Phaedrus* 57/244a). Both the fact that he puts value on madness as such, and the condition attached that it needs to be divinely given, contrast with our modern ideal of the person. Not madness, but the exercise of reason is part of the modern ideal. As Berlin puts it, I wish 'to be moved by reasons', and if I say that I wish to direct myself, it is because I insist that 'I am rational, and that it is my reason that distinguishes me as a human being from the rest of the world' (Berlin 1997 (1958): 203). Furthermore, the ideal of self-determination seems to be – *prima facie* at least – in contradiction to the requirement that anything at all should be divinely given to us. So we want to know the following. Firstly, we want to explore what it could mean to say that the ideal person is the mad person (according to Plato and according to the culture of his time – these views *may* be different). Secondly, we want to find out whether this ideal or these ideals contrast with ideals in other writings of Plato, such as the ideal in his story of the charioteer managing the horses (see beginning of this chapter). Thirdly, we want to find out whether, and if so, to what extent, this ideal or these ideals contrast with the modern ideal of the person as an autonomous individual.

Within the *Phaedrus*, there is an apparent contradiction between the stress on reason and self-control in the image of the charioteer and the horses, on the one hand, and Plato's praise of madness, divinely given, on the other hand. Plato ascribes all kinds of good qualities to the white horse, representing the rational impulse, and the charioteer (at least so it is commonly interpreted) represents reason. At first sight, this is (just) an argument about self-control. If people are masters of themselves and are orderly, they will pass their life in happiness. In contrast to this, his claims about madness leave the modern reader wondering whether Plato meant this seriously (see for example Jowett 1953: 122). However, it can be shown that if we understand Plato's claims in the context of this *whole* ideal of the person, they do make sense and are even a necessary element of it. Therefore, we need to take his claims about madness seriously.

Plato makes a distinction between madness that is evil and madness that is a divine gift, the source of the greatest blessings granted by gods to man. He distinguishes between (following Dodds (Dodds 1951: 65;

265b)) prophetic madness, ritual madness, madness brought about by the Muses, and – according to Plato the most important one – the madness of 'the lover of the Beautiful' (249d). To understand this, we have to go back to the story of the horses, since this story is, according to Plato, the proof that the madness of love is the greatest of heaven's blessings. This puts Plato's claims about madness in the right context and makes them more intelligible.

As we saw, Plato compares the soul to a pair of winged horses, one of noble breed, the other ignoble (246a–b). Now the story tells us that our souls had seen the truth but now only few men retain an adequate memory of it. The philosopher (the ideal person), however, is able to recollect the idea of the truth, the good, the beautiful, etc.[11] The one who employs the memories rightly becomes truly perfect: 'But, as he forgets earthly interests and is rapt in the divine, the vulgar deem him mad, and rebuke him; they do not see that he is inspired' (249d). So it is not only that other people call the philosopher mad, he also *is* mad, according to Plato, 'rapt in the divine'. This is the fourth kind of madness, a madness which is imputed to 'him, who, when he sees the beauty of earth, is transported with the recollection of the true beauty; he would like to fly away, but he cannot' (249d). Indeed, it hasn't been said that being a lover of the beautiful, or having vision, is necessarily easy or pleasant. If a soul recalls things from 'the other world', if someone remembers, recalls 'the good' or 'the beautiful', he may experience this as a struggle – as we will soon see.

The ones who are 'rapt in amazement' are now confronted with 'earthly things'. The true lover will then recognise 'heavenly' beauty in the earthly things, in other people for example. This contrasts with the attitude of the non-lover: 'He who is not newly initiated or who has become corrupted, does not easily rise out of this world to the sight of true beauty in the other, when he contemplates her earthly namesake, and instead of being awed at the sight of her, he is given over to pleasure' (250e). He is a 'wanton' (251a), in Plato's terminology as well as in Frankfurt's. So although we have all once seen the truth, the good, the beautiful, we may (1) not recollect it or (2) recollect it but then become corrupted. The lover, on the contrary, 'is amazed when he sees anyone having a godlike face or form, which is the expression of divine beauty; and at first a shudder runs through him, and again the old awe steals over him; then, looking upon the face of his beloved as of a god he reverences him, and if he were not afraid of being thought a downright madman, he would sacrifice to his beloved as to the image of a god' (251a). We can infer from this that madness is essential to what it is to be an ideal person, a philosopher, a

lover of the good. It may be that the lover 'has forgotten mother and brethren and companions, and he thinks nothing of the neglect and loss of his property' (252a). But he is, according to Plato, the winged one, the ideal person.

I conclude that Plato's ideal of a person is marked by a stress on self-control. However, to reach this state the person has to have a vision of the good (the beautiful, the truth) and the one who has this vision is someone who can be regarded as 'mad' in the precise sense Plato explains. There may be other beneficial forms of madness, as well as non-beneficial or even 'evil' forms, but *this* madness is about the one who is in a state of 'rapture' at the sight or rather recognition of the divine *in this world*. Only because of this 'madness' is it possible for the charioteer (reason, love?) to manage the horses, in other words, to reach inner harmony and self-control, which are seen by Plato as the virtues of the ideal person. I can conclude that madness is an essential part of Plato's model of the soul and the ideal of the person I inferred from that model.

Dionysian madness and madness in ancient Greek culture

One of the other forms of madness Plato mentions is Dionysian madness. Does this kind of madness have the status of an ideal too? Does it have this status in Plato's work? And does it have this status in ancient Greek culture as a whole? It is helpful to discuss these questions to arrive at a more adequate and comprehensive picture of the Platonic ideal of the person, which will allow me further to explore and discuss the compatibility of this ideal with the modern ideal of autonomy.

The issue of Dionysian madness arises because we are confronted with what at first sight appears to be an anomaly or contradiction in the *Phaedrus*. On the one hand, the metaphor of the charioteer and his two horses conveys the ideal of self-control. My argument that vision of the good is an essential part of this ideal does not change the fact that it is an ideal of self-control. On the other hand, we find in the same piece of writing Plato's praise of madness, which includes Dionysian madness. According to Plato, madness is not always bad: 'It might be so if madness were simply an evil; but there is also a madness which is a divine gift, and the source of the chiefest blessings granted to men' (244a). With regard to what Dodds calls 'ritual madness' (Dodds 1951: 65), of which Dionysian ritual is an example, Plato refers to 'religious rites' which can have a beneficial effect:

> And learning thence the use of purification and mysteries, it has sheltered from evil, future as well as present, the man who has some

> part in this gift, and has afforded a release from his present calamity to one who is truly possessed, and duly out of his mind. (244e–245a)

Having read this we wonder whether 'being possessed' and 'being out of one's mind' is compatible – as an ideal – with the ideal contained in the story of the charioteer and the horses. The latter is, as I have shown, in part compatible with the modern ideal of autonomy. But 'being possessed' and 'being out of one's mind' seems to contrast sharply with being in a condition of autonomy. If I'm possessed by something, it's not the case that 'I rule me and no one else rules I' (I will say more about this below). Furthermore, the 'ideal' of ritual madness seems to contrast with Plato's ideal of self-control contained in the story of the charioteer and the horses. It could be argued that Plato only recommends ritual madness to certain people. But even then his praise of divinely given madness in general, including prophetic madness and the madness of those who are possessed by the Muses, is still problematic. If it is ideal that something is divinely given to us, how can our charioteer in the story be supposed to remain in control of the horses? Descriptions of madness seem to suggest a complete loss of control. Phenomena such as possession, etc. seem to make the charioteer redundant: the horses, it could be argued, are taken over by a divine charioteer, a daemon, or the black horse is given complete freedom, or, alternatively, we could say that both horses are 'bewitched'. Whatever image we use, the result is the polar opposite of self-control. Moreover, this ritual madness does not only contrast with Plato's ideal of the person contained in the story of the charioteer and the horses, it also contrasts with the modern ideal of autonomy, and this latter for two reasons. Firstly, there is a lack of 'inner' autonomy. This is sufficiently clear already since there is obviously a loss of self-control. But, secondly, as I suggested above, there is also a lack of 'outer' autonomy: If I am taken over by a god or a daemon, I can not, or no longer, say 'I rule me', since it is not *me* but this 'other' who rules me, and this destroys any possibility of autonomy in the sense of independence. So we do not only wonder why Plato praises ritual madness because we know his ideal in the *Phaedrus* as one of self-control, we also perceive a gap between any possible 'ideal' of ritual madness and the modern ideal of autonomy. Indeed, we wonder whether, given the two ideals already discussed, ritual madness can be (part of) an ideal at all. To reach a final conclusion on this matter, however, I need to inquire further into the nature of ritual madness, and Dionysian madness in particular. I choose to

focus on *Dionysian* madness in particular because there are sources available to say something about it and because it sheds light on the characteristics of ritual madness, and the relation between those characteristics and autonomy.

In Euripides' play *Bacchae* the story is told of a man, Pentheus, who refuses to recognise the divinity of Dionysus and is therefore punished: he is torn apart and killed by the women, the Bacchae, led by his own mother. This brief plot tells us something about the danger of ritual madness. Psychologically, being out of your mind may release destructive forces. It may destroy yourself or others, whether these others are involved in the 'ritual' or not. On a social level, it may mean that a scapegoat is ritually killed. René Girard (1999) has argued that it was the achievement of Christianity to put – at least in theory[12] – an end to this scapegoat-ritual (arising from what he calls 'mimetic' desire) and instead to defend the victim: 'ethics' against 'mimesis', to use his terms. However, it would distract from the purposes of this chapter to explain and discuss this argument fully. The main point made here is that ritual madness is dangerous and potentially lethal.

Given this short analysis of the danger of Dionysian madness, it is clear that it cannot be a condition of autonomy. On the contrary, instead of self-rule, it is a condition of *being ruled* by one's 'inner' psychological forces and/or by 'outer' social forces. Furthermore, it is hard to see how Dionysian madness ever can be an ideal of the person at all.

However, there may be beneficial aspects of Dionysian madness. Following Dodds' account of the matter (Dodds 1951: 76), it can be argued that the ritual provided some positive functions too. Firstly, psychologically it means for the person involved an outlet for irrational impulses, which may 'release' the person in a certain sense. This is what Plato means by 'release'. Of course, as argued earlier, this may have destructive effects, both for the person involved as for others. But the 'good side' here is perhaps a feeling of 'liberation' on the part of the person 'letting go'. Dodds also suggests a 'release' from having to be a certain individual or having to play a certain role in society. He writes about Dionysus:

> he is Lusios, 'the Liberator' – the god who … enables you for a short time to *stop being yourself*, and thereby sets you free. That was, I think, the main secret of his appeal to the Archaic Age: not only because life in that age was often a thing to escape from, but more specifically because the individual, as the modern world knows him, began in that age to emerge for the first time from the old solidarity

> of the family, and found the unfamiliar burden of individual responsibility hard to bear. Dionysus could lift it from him. For Dionysus was the Master of Magical Illusions, … 'Dionysus leads people on to behave madly' – which could mean anything from 'letting yourself go' to becoming 'possessed.' The aim of this cult was *ecstasis* – which again could mean anything from 'taking you out of yourself' to a profound alteration of personality. And its psychological function was to satisfy and relieve the impulse to reject responsibility…. (Dodds 1951 76–7)

Secondly, socially, the ritual erases differences between people and builds a kind of bond between people. Whether you were a slave or a free man, a woman or a man, rich or poor, in the Dionysian ritual there is equality in joy. Thirdly, this social bond and union, it could be argued, is only part of a more general 'mystical' union that brings about harmony. Nietzsche writes:

> Now the slave is a free man [….] Now, with the gospel of world harmony, each man feels himself not only united, reconciled, and at one with his neighbour, but one with him, as if the veil of Maya had been rent and now hung in rags before the mysterious primal Oneness. Singing and dancing, man expresses himself as a member of a higher community [….] His gestures express enchantment. Just as the animals now speak, and the earth yields up milk and honey, he now gives voice to supernatural sounds: he feels like a god [….] Man is no longer an artist, he has become a work of art [….] (Nietzsche 1993 (1872): 17–18)

In *The Birth of Tragedy*, Nietzsche clearly views Dionysian madness as an ideal. Moreover, the psychological and social functions described by Dodds point to beneficial functions of the ritual. But, as we have seen, there is also a 'dark' side to the ritual; these voices may present a too-one-sided view of Dionysian madness. Moreover, and this is the most important point, what I have called 'beneficial' functions of Dionysian madness may be beneficial in the way described above, but they are not beneficial to the achievement of autonomy for at least the following three reasons.

Firstly, if you're out of yourself, there is no way you can (simultaneously) rule yourself. Furthermore, if anything like 'a profound alteration of personality' is possible at all, within the ideal of autonomy this would be seen as having to be the result of self-direction and

radical self-evaluation (I will say more about autonomy and identity later), and certainly not the result of possession by the 'Master of Magical Illusions.' And the psychological function of satisfying and relieving 'the impulse to reject responsibility' (Dodds 1951: 76–7) is in sharp contrast with the modern wish to be 'a thinking, willing, active being, bearing responsibility for my choices and able to explain them by reference to my own ideas and purposes' (Berlin 1997 (1958): 203). It may be that the ideal of being not responsible is an ideal according to some people, but it is certainly not the dominant modern ideal of *autonomy*.

Secondly, erasing the (social) difference between people may be pleasant (especially for the ones who normally have a low social status) and some might interpret Marxist doctrine as encompassing such a view as an ideal, but in any case it is unclear to me why this would enhance the autonomy of the persons involved.

Thirdly, Nietzsche's description of Dionysian mystical union in terms of 'feeling like a god' may appeal to persons who interpret autonomy as 'doing what you want', but (1) I have rejected this interpretation already, and (2) the bulk of the content of the passage quoted above stresses being part of a whole, of a oneness, a mystical union which erases the boundaries between persons and therefore the possibility of autonomy. If I can no longer make a difference between 'me' and 'not-me', and between 'I' and 'you', it doesn't make sense to say 'I wish to rule myself' and 'I wish that no one rules I'. The condition of mystical union is incompatible with autonomy.

Dionysian madness, then, is not an ideal of autonomy. Moreover, it can hardly be called an ideal at all. On balance, it seems not appropriate to call it an ideal of the person. But was it an ideal of the person in ancient Greece? Arguably not as such; it seems very unlikely that somebody can seriously propose this as an ideal of the person, in the sense that the ideal person would be constantly 'out of his mind', 'possessed', etc. Rather, it may be claimed that it can be *sometimes* ideal, for example, to refer to Plato, if 'relief' is beneficial. But this does not make it an ideal of the person, not according to us and not according to Plato.

To conclude, this excursion into the nature of Dionysian madness makes Plato's claim about Dionysian madness more intelligible (considering the beneficial aspects pointed to above), but also requires me to conclude that this is an aspect of Plato's thought that cannot be reconciled with the modern ideal of autonomy.

Even if we accept this conclusion, there remains a problem with the coherence of Plato's thought. I want to discuss this problem, since later in this book (Chapter 4) I will make use of Plato's concept of 'the good' as 'the one'. Is such a notion not 'mystical', and therefore (as I argued above) in contrast with autonomy? Consider the following three notions of oneness at work in my discussion: (1) Platonic (universal) 'oneness', (2) Nietzsche's concept of a 'primal Oneness', and (3) the oneness which is the unity of the person. How are the first two notions related? If we accept Nietzsche's description of what we could call the 'mystical' aspect of Dionysian madness as a valid interpretation of these particular ancient Greek phenomena, we may wonder in what way the concept of a 'mysterious primal Oneness' is related to the Platonic 'oneness', the oneness that is part of the ideal of the person in the *Republic* (see Section 2.2.2 below). They are at least very different in regard to the way to reach 'oneness'. To reach the Platonic 'oneness' it is recommended to use your rational faculties to study the sciences and make abstractions from the 'many' and the 'particular' to arrive at an idea of the 'one' and the 'universal'. The Dionysian 'Oneness' Nietzsche mentions, however, has little to do with what is rational. The ecstasy of Dionysian madness is a going 'out of oneself'. But can the Platonic ascent to 'the good' not be understood as a going 'out of oneself' to the good? Plato's ideal of the person discussed so far does not allow for this. Instead of a mystical unity between 'me' and 'the one', the Platonic ideal of the autonomous person includes *reference* to 'the one' (the good) but there is a 'going out of oneself' only in the sense of a contemplation of 'the one', the good.

I conclude that Plato's ideal of the person is not compatible with Dionysian madness, given the full consequences of this phenomenon if it were to be made into an ideal of the person. However, this conclusion still leaves us with some dissatisfaction. We know that Plato *does* praise divinely given madness, and that he – in the *Phaedrus* and elsewhere – *does* refer to religious elements such as gods and daemons. Is there perhaps a way I can retain (in my picture of Plato's ideal of the person) some aspects of this reference to something 'outside' the person which is not 'the one' and still influences me? If I widen my attention from Plato to ancient Greek culture in general, there are at least two aspects relevant to Plato's ideal of a person I want to discuss, since they are problematic in relation to the ideal of the autonomous person. Both aspects have to do with the belief that we can't control everything.

The first aspect has to do with the belief that we cannot control everything because there exists a certain 'order' in the universe; therefore, human judgement is constrained:

> Life is so vast, complex, uncertain, that we delude ourselves if we think that we can control it; human judgement is fallible, over-reliance on it leads to hybris, and that always ends in disaster. Many things may be inexplicable, but life is not random; the gods do exist and their laws do work. If we think that there are no laws, that we can take each thing as it comes, neglect the restraints and sin intelligently, we are only deceiving ourselves. (Kitto 1961: 186)

The deterministic picture sketched here seems to be completely contrary to the ideal of the autonomous person. If there is a pre-existing order, including gods and their laws, how is autonomy possible? It seems that the ancient Greek world view and the modern world view are *totally* irreconcilable.

A second aspect of ancient Greek culture is equally difficult to understand from within the modern ideal of autonomy. Again, it is not so much the idea of 'outer' autonomy that is challenged but 'inner' autonomy. Here there is a loss of control because of passion 'possessing me'. Dodds writes:

> The Greek had always felt the experience of passion as something mysterious and frightening, the experience of a force that was in him, possessing him, rather than possessed by him. The very word *pathos* testifies to that: like its Latin equivalent *passio*, it means something that 'happens to' a man, something of which he is the passive victim. Aristotle compares the man in a state of passion to men asleep, insane, or drunk... the men of the Archaic Age interpreted such experience in religious terms, [...] as a communication of *menos* [might, force, strength, fierceness, spirit, passion], or as the direct working of a daemon who uses the human mind and body as his instrument. (Dodds 1951: 185)

Note that the same experience (that of passion overwhelming me) can be interpreted as a problem of 'outer' autonomy or 'inner autonomy'. The 'outer' autonomy problem with regard to a daemon possessing me has already been discussed in the context of the contrast between Dionysian madness and Platonic reason. Now it could be argued that, whatever the problem for Plato's ideal of the person, we moderns do

not need to bother with daemons and gods. In that case, let's interpret the problem of 'passion' not in terms of possession of a daemon but as an 'inner' overwhelming emotion. This would be a way of putting it to modern readers who may be sceptical regarding the use of religious or supernatural concepts. But this is still a problem for the modern ideal of autonomy. The idea is that there could be an emotion so overwhelming that it 'takes over', that I lose control. Surely this is not something we want if we wish to be autonomous. How could this be part of an ideal of the autonomous person? It may be objected that we could interpret 'passion' as a first-order desire we are able to control. But the reason why the ancient Greeks were so afraid of 'passion', or interpreted it as the work of a daemon, was precisely the feeling that 'passion' *happens*. This essential aspect of is lost if 'passion' is interpreted as a first-order desire I can (decide to) control. The reason why the ancient Greeks were afraid of it, or interpreted it as the work of a daemon, was precisely the feeling that the 'passion' *happens* to me, that it is *alien* to me, that it comes from the world 'outside' me, that it is *not* me. Within modern thinking generally averse to using religious terms there is no way of incorporating this particular human experience into the ideal of the autonomous person. This is a limitation of the modern ideal of autonomy, since it is not able to 'capture' such experiences, but it is a necessary limitation if we want to uphold the concept of autonomy itself. The function of a concept is to draw distinctions (within our thinking and/or our experience of the world), and this means that some experiences and ideas have to be excluded. Within the framework of the ideal of autonomy, and in relation to the experience(s) discussed, we have to compare the self here with a kind of citadel, as is suggested by Berlin (Berlin 1997 (1958): 207). (I could compare 'passion' here with either a visiting stranger or an alien force taking me over. If a stranger enters the citadel, I can let him in or not. If I let him in, I necessarily make him *my* guest. If the citadel is taken over by alien forces, however, the autonomy of the citadel is lost.) There may be other ideals worth discussing, but I have to limit myself here to the ideal of autonomy.

2.2.2. Book VII of the *Republic*

So far I have discussed Plato's ideals of the person in the *Phaedrus* and the various forms of madness related to (some of) these ideals. Indeed, although I set out to discover one ideal, there seem to be various related ideals, often connected with a particular metaphor or cultural practice. For now, I have to conclude that these ideals are all in some

respects incompatible with the modern ideal, and that some are more distant from it than others (consider for example the metaphor of the charioteer and the horses versus Dionysian madness). To further develop the picture I am sketching of Plato's ideal(s) of the person, however, I will consider Book VII of the *Republic*. In particular, this discussion brings out the Platonic idea of *vision*.

If combined with the rest of Book VII, the famous story of the cave gives us the following picture of the relation between vision, the good, and the ideal(s) of a person according to Plato. Firstly, the ideal 'to see the Idea of the good' is only reached at the end of the journey upwards. It takes time and effort to reach this ideal. Plato says that 'in the world of knowledge, the Idea of the good appears last of all, and is seen only with an effort' (517b). Secondly, the ideal of 'vision of the good' is not contrary to rationality. Rather, according to Plato, one who wants to act rationally must have this vision. The Idea of the good is 'the power upon which he who would act rationally either in public or private life must have his eyes fixed' (517c). So proper rational action supposes attention to the Idea of the good. Thirdly, if it is possible to look at the good, it is also possible to *not* look at it. The person may not look 'upwards' but 'downwards'. Then this person is evil (in the sense of lacking (knowledge of) the good).[13] In that case, the person is not 'lost'. We all have 'the power and capacity of learning' (518c) and it is possible to move the soul, to turn it. We do not need to 'implant the faculty of sight, for that exists already, but will set it straight when it has been turned in the wrong direction, and is looking away from the truth' (518d). If the evil persons would be 'turned in the opposite direction, the very same faculty in them would have seen the truth as keenly as they see what their eyes are turned to now' (519b). We all have the faculty of vision; the (moral) problem is a question of *direction*.

I will turn to this question of direction in the next chapters. For now, note that this notion of vision appears to be alien to the modern ideal of autonomy. According to the modern ideal, the question is whether or not I rule myself, not whether or not my soul is 'turned in the right direction', i.e. the direction of the good. The idea of there being a soul or the idea of the good is, by itself, not widely and readily accepted, let alone that it could be widely and readily accepted as being an ideal of autonomy.

The same can be said about the methods Plato recommends to bring people 'from darkness to light' (521c), that is, the methods to make people turn to the (Idea of the) good. What does *that* have to do with

autonomy? What does autonomy have to do with 'the contemplation of true being' (525a), with knowledge of 'the unseen' that can 'make the soul look upwards' (529d), with 'the Craftsman' who framed 'the things in heaven' (530a), with 'the end of the visible' (532a–b)? If Plato writes that to arrive at the Idea of the good is a matter of 'elevating the highest principle in the soul to the contemplation of that which is best in existence' (532e), why is this necessary or sufficient for achieving *autonomy*? It may be an ideal of the person to study philosophy (535c) and to be a lover of 'learning or listening or inquiring' (535d) rather than 'wallowing like a swinish beast in the mire of ignorance' (535e). But it is unclear, at present, what all this has to do with autonomy.

Note that the ideal of the person in Book VII, as summarised above, was never meant to be an ideal for everyone. Plato reserves it for the Guardians in his ideal state. However, this does not stop it from being an ideal of the person for Plato. It is clear enough that he prefers the qualities of the Guardians (and the philosophers) to those of the rest of the people. But this only makes his ideal clearer: persons 'must raise the eye of the soul to the universal light which lightens all things, and behold the absolute good' (540a). So far, however, it is unclear how such an ideal could ever be reconciled with the modern ideal of the autonomous person.

2.3. Augustine's ideal of the person

Having articulated Plato's ideal(s) of the person, and having argued that it seems to be irreconcilable with the modern ideal of autonomy, I turn now to Augustine with the same aim. More precisely, I want to (1) articulate Augustine's ideal of the person on the basis of his discussion of freedom in *On Free Choice of the Will* and his self-understanding in the *Confessions*, and (2) inquire whether there is a chance that – however interesting and attractive it may be in other respects – this ideal is compatible with our modern ideal of the autonomous person.

2.3.1. *On Free Choice of the Will*

Although *On Free Choice of the Will* mainly deals with the question of free will and the origin of evil, I will argue that in his discussion Augustine expresses a certain ideal of the person. In particular, and like Plato, Augustine praises the capacity for self-control: 'whatever it is that sets man above beast […] if it controls and commands whatever else man consists of, then man is ordered in the highest degree' (I,18[14]). Augustine clearly favours the condition of self-control and inner order.

There is apparently much in common here with the modern ideal of the autonomous person. Dilman's interpretation suggests this too. He interprets Augustine's phrase 'When I willed or did not will something, I was wholly certain that it was not someone other than I who willed or did not will it'[15] as meaning that Augustine 'had no doubt that he was the author of his decision and action, that he had himself formed the intention in his action. That is, his decision, intentions, actions were "his"' (Dilman 1999: 73). Dilman understands this 'his' in terms of what we have called 'outer' autonomy as well as 'inner' autonomy. Firstly, he presumes that Augustine was not just following the opinions or convictions of his comrades. This is autonomy in the sense of being independent from 'opinion'. Secondly, Dilman discusses the 'inner' sense when he writes that Augustine 'was unable to put himself wholly behind what he willed. For he was divided in himself. So in part of himself he remained unwilling. […] He could not put his whole self behind his will – whole heartedly. He remained disunited in himself until his conversion' (Dilman 1999: 74). This is the opposite of autonomy. But what would 'inner' autonomy be, then? When was Augustine autonomous? It could be argued: when Augustine 'came together in himself'.

> That is where one is clear and has no hesitation about what one is to do there is no effort of will to be made and so no 'willing' in that sense. There is then no question for one about what one must do. If anyone else questions it and suggests an alternative, the natural answer is: 'I have to do it.' It is in this sense that the action one is ready to embark on presents itself to one as a *necessity*. There is no division then between what one feels one ought to do and what one wants to do: one wants to do with one's whole being what one knows and feels one ought to do. (Dilman 1999: 73–4)

So far, so good. This is the picture of 'inner' autonomy sketched so far. I came to the same conclusion in my discussion of Plato: the condition of 'inner' autonomy is a condition of order and harmony in the soul. In Frankfurt's terms: we are able to identify with one of our desires. The first-order desire is not just 'regulated' by the higher-order desire. This suggests control and command, and therefore division. Rather, the person is in a state of harmony, undivided within himself.

Note that the idea of *necessity* in this context is not necessarily in contradiction with autonomy. Frankfurt himself has proposed a concept of 'volitional necessity' to account for this (Frankfurt 1988). I will say more about this in Part II (Chapter 6, in particular Sections 6.2.2. and 6.4.1.).

So far, it appears that Augustine's ideal of the person corresponds well with the modern ideal of autonomy, including Frankfurt's model. However, there are also important differences between the two. To start with, there is an important difference between Augustine and Frankfurt. Whereas Frankfurt mainly stresses the volitional aspect of this self-control,[16] Augustine puts more or at least as much emphasis on the faculty of reason. When are we well ordered within ourselves? 'When reason is master of the emotions, a man may be said to be well ordered' (I, 18). The key difference between this ideal of self-control and self-mastery and the modern one, however, is that Augustine legitimates the ideal by reference to 'the eternal law', to some kind of pre-existing order in the universe: 'when reason, whether mind or spirit, rules the irrational emotions, then there exists in man the very mastery which the law we know to be eternal prescribes' (I, 18–19). By looking at what Augustine takes this eternal law to prescribe, we get a good idea about his ideal of the person: 'do you think there is anything more excellent than a rational and wise mind?' (I, 21). Augustine favours 'the mind that possesses virtue and is in control' (I, 22). According to Augustine, being in control is itself a virtue, if not (one of) the most important virtue(s), since if we do not exercise this virtue 'the reign of lust rages tyrannically and distracts the life and whole spirit of man with many conflicting storms of terror, desire, anxiety, empty and false happiness, torture' (I, 22). So his argument is not just that we need to exercise self-control because it is prescribed by the eternal law; it is also wise to do it since otherwise we punish ourselves.

Augustine mentions the virtues of prudence, fortitude, temperance, and justice. Again, temperance is a virtue Augustine values very much. He defines it as 'a quality which checks and controls the desire for those things that it is base to desire' (I, 25). It is 'the virtue which restrains lust' (I, 26). Note that, according to Augustine, self-control does not stand on its own, but necessarily includes a *judgement* about desires: some are better than others. Augustine calls 'those things that it is base to desire' the lust after temporal things. (This may be read as an expression of (part of) what I have called 'doing what you want'; however, I do not want to identify the two too strongly, since 'doing what you want' as such does not depend on the assumption of a distinction between 'temporal' and 'eternal' things. In other words, we don't need the Augustinian metaphysics to critique 'doing what you want' as being an inadequate conception of the ideal of autonomy.) Furthermore, in the context of his discussion of what 'the temporal law' (as opposed to the eternal one) commands, he mentions what we

could call the ideal of political freedom. The temporal law commands 'freedom – not, indeed, true freedom, which is reserved for those who are happy and who abide by eternal law; rather, I am speaking now of that freedom which men who have no masters think they possess, and which men who wish to be free of human masters desire' (I, 31). In other words, Augustine says that we might be free in the sense of having no master, being not a slave. This is (part of) what it is to have political freedom and freedom of action, but, according to Augustine, this is not true freedom. He claims that the only true freedom is the one that rewards those who follow the 'eternal law'. If we become a 'citadel of mastery' (I, 33), we do not only follow the eternal law, but we are also rewarded for it in terms of freedom and happiness.

Augustine's ideal of the person has very much in common with the Platonic one. Self-direction towards 'the one' is held as necessary to reach inner order and harmony. For Augustine, however, 'the one' is explicitly understood as being divine or at least of direct divine origin. But the basic divide between the temporal and the eternal, the many and the one remains in force:

> All sins are included under this one class: when someone is turned away from divine things that are truly everlasting, towards things that change and are uncertain. These things have been rightly placed in their own order and complete the universe through their own peculiar beauty; but nevertheless, it is characteristic of the perverse and disordered spirit to be a slave to the pursuit of the things which divine order and law have prescribed should follow its own binding. (I, 34)

Again we see that the ideal of the person is often defined by expressing what it is not to be an ideal person. If it is sinful to turn away from divine things, it is good to turn towards them. The ideal person is the one who is ordered properly because of this self-direction towards a divine order.

Whether or not we turn away from eternal things – in Augustine's words, whether or not we sin or commit evil – is a matter of the free choice of our will. According to the view of autonomy developed in this book so far, this would imply that the ideal person is not only autonomous in the sense of being in a state or *condition* of inner autonomy (well ordered, in harmony, self-controlled); this ideal of the person also contains autonomy as a *capacity*. I have a will that is able to freely choose to turn to eternal things or not.

However, the compatibilities discussed so far cannot conceal the fundamental problem faced by any attempt at reconciliation between ancient and modern ideals of the person. We wish to be autonomous persons but most of us are not prepared to accept Augustine's metaphysics. Why, if I want to rule myself, do I need any concept such as 'the one', 'things eternal', or 'God'? Do these terms not, rather than having anything to do with autonomy, point to a relationship of *dependence* which excludes my autonomy? On the one hand, Augustine suggests that it is up to me what I do with my free will (which is given to me by God). This view is compatible with modern autonomy. It is not because God has given me a capacity (the capacity of free will) that I lack freedom in relation to what I do with that capacity. This is also true for autonomy. I am able to rule myself, also if this capacity to rule myself is given to me. To use Wolf's words: I am not '*metaphysically* responsible' for myself since I did not create myself but I am '*morally* responsible' since I am 'able to understand and appreciate right and wrong' (Wolf 1989 (1988): 147). (I will say more about Wolf's view in the next chapter (Section 3.4.).) However, on the other hand, as his discussion of the problem of the origin of evil shows, Augustine argues that the (moral) responsibility for the person's deeds is shared between the person and God. Although according to Augustine I am fully responsible for my evil deeds – Augustine attributes these 'only to man's will' (III, 137) – I am, in Augustine's view, not (entirely) responsible for my good deeds. Augustine's view on responsibility is highly asymmetric. Whereas God is not responsible for our evil deeds, he seems to claim merit for our good deeds. If 'all good proceeds from God' (II, 48) means that God is responsible for our good deeds, it seems that there is not much autonomy left for us. One could object that responsibility is not necessarily fully exclusive: we may 'share' responsibility for our good deeds with God. However, the problem then is that (1) our autonomy is still questionable since then we are for our good deeds at least in part dependent on God and (2) it is difficult to see why our *evil* deeds are not a matter of shared responsibility with God. To conclude, the only viable version of this ideal of the person is to say that God has given us the capacity to choose between good and evil, but that then it is up to us to decide. But this view still suffers from relying on complex metaphysical commitments. Why should terms such as 'God' be part of the ideal of autonomy at all? It seems that I am very well able to say what I mean with 'I wish to be autonomous' without reference to God.

Although I have indicated some compatibilities with the modern ideal of the autonomous person, I have shown that there are good reasons why a full reconciliation between ancient and modern ideals of the person will be very difficult to achieve. Augustine's discussion of the origin of evil bears this out most clearly. Certainly, Augustine's belief that free will is a gift of God can be read as a way of praising the person's capacity for autonomy. However, according to Augustine's ideal of the person we want to turn towards 'things eternal' – God, or, in Platonic terms, the good. As was the case with Plato's ideal, it is difficult to see how this can be part of an ideal of autonomy.

This difficulty only grows if we consider the Augustinian concept of grace. If we 'cannot do right, follow goodness, without God's grace' (Dilman 1999: 73), as Dilman interprets Augustine, then are we still autonomous at all? How can this be an ideal of the autonomous person? Why would I wish to be dependent on God to do what is right or to follow goodness? Why, if I want to reach inner autonomy in the sense of inner harmony, would I want to be dependent on God? Do I really rule myself if I am dependent on grace to reach inner harmony?

To do justice to Augustine, it is important to not confuse grace with pure passivity. God is not seen as the one who does everything. We still have to make an effort. In Dilman's words:

> We must neither take credit for it, nor just wait for it to fall into our lap. We have to put ourselves out for others, struggle with pride and temptation, and keep our souls turned towards God and open to Him. This is what I understand Augustine to be saying when he says that 'man cannot rise of his own will'. (Dilman 1999: 81)

Note that the language used here suggests a very different way of thinking about persons than the modern way. We will discuss the terms 'turning' and 'opening' in the next section when we look at Augustine's *Confessions*.

The Augustinian ideal of the person articulated so far is that of a person who is able to exercise self-control and see 'things eternal' with the necessary help of God. If we turn ourselves towards God and open ourselves to Him, we will be good, innerly ordered, excellent persons. If we reach this ideal, we can't 'take the full credit for it', but we will have gone through a struggle that requires a lot of our own effort. This ideal seems to be incompatible with modern thinking about autonomy, and so far it is difficult to see what argument could be offered for the necessity of terms such as 'good' or

'God' for the ideal of autonomy to be coherent. Augustine's concept of grace is particularly problematic in relation to efforts to reconcile 'ancient' and 'modern'.

This conclusion does not mean that my discussion has not yielded interesting material relevant to the question of autonomy, and I would like to continue my engagement with Augustine's thoughts by looking at his *Confessions*. This will allow me to present a more comprehensive picture of his ideal of the person, and further to show in what ways it contrasts with the modern ideal of autonomy.

2.3.2. *Confessions*

The *Confessions* shows us at least three problems in regard to attempts to reconcile Augustine's ideal of the person with the modern ideal of autonomy. Firstly, Augustine argues that since we have free will we can choose between good and evil. The problem is on what grounds such a choice can be made, and why we are responsible if we do evil but not if we do good. Secondly, Augustine's ideal of 'inner harmony' is, by itself, compatible with the modern ideal of autonomy, in particular 'inner autonomy'. But the fundamental difference between these views is that, according to Augustine, I need God's grace to reach the ideal. Thirdly, Augustine's view contains elements such as love, dependence, and receiving. It is unclear how such elements could be compatible with the modern ideal of autonomy as self-rule.

The first problem appears to be an 'internal' difficulty but will turn out to be very relevant for my further discussion. The others are more straightforward problems that emerge if we would want to reconcile his ideas with the modern ideal of autonomy. The first problem already figured in the discussion above; the third one is new. I will now discuss each of them in turn.

1. Free will and the origin of evil

The origin of good is God, according to Augustine. But what is the origin of evil? He claims that 'we do evil because we do so of our own free will…' (VII.3/136). His argument goes as follows:

> I knew that I had a will [….] When I chose to do something or not to do it, I was quite certain that it was my own self, and not some other person, who made this act of will, so that I was on the point of understanding that herein lay the cause of my sin. If I did anything against my will, it seemed to me to be something which happened to me rather than something which I did […] (VII.3/136)

The problem he still has, however, is that, if we indeed have a free will, a will which enables us to do good or wrong, why do we choose to do wrong? If it is not something which *happens* to me but is an act of my *will*, why do I do wrong things? He enquires into the nature of evil. 'How, then, do I come to possess a will that can choose to do wrong and refuse to do good […] ? Who put this will into me? Who sowed this seed of bitterness in me…? It was the devil who put it there […]' (VII.3/136–7). But this, he realises, is no answer to the question. If the devil is the answer to the origin of evil, we may still wonder why the devil came to be wicked. 'How did he come to possess the wicked will which made him a devil, when the Creator, who is entirely good, made him a good angel and nothing else?' (VII.3/137).

Augustine wants to maintain that the freedom of the will includes the freedom to choose between good and evil. But the difficulty is that if there is no ground for choosing evil, there is no reason to extend freedom to the choice between good and evil. If we don't have a ground for choosing evil, then why is there still any choice at all? This problem will return regularly in the next chapters, since it is a problem not only for Augustine, but, I will argue, also for anyone who wants to hold on to the modern ideal of the autonomous person.

2. Grace

Augustine's story of his life is a story of a struggle for inner autonomy. First, Augustine himself does not appear to be fully autonomous:

> I was held fast, not in fetters clamped upon me by another, but by my own will, which had the strength of iron chains. […] So these two wills within me, one old, one new, one the servant of the flesh, the other of the spirit, were in conflict and between them they tore my soul apart. (VIII.5/164)

This is not the opposite of outer autonomy, as Augustine says that the fetters are not clamped upon him by another person, but the opposite of inner autonomy. Augustine's metaphysics include such things as 'flesh' and 'spirit', which may be inconsistent with ours, but we can easily interpret this in terms of conflicting desires. To use the Platonic image: the two horses are not running in the same direction, and there is a lack of harmony between the charioteer and the horses. 'My inner self was a house divided against itself,' Augustine writes (VIII.8/170). This is anything but an ideal of the person. But if a person is in this condition, how can he get out of it?

How can a condition of non-autonomy or anti-autonomy become one of autonomy? If autonomy is an ideal of the person, we want to know how this ideal can be reached.

As I have argued already, Augustine's answer is that we can reach this ideal by our own efforts in combination with the grace of God. Augustine's appeal to grace is for him the only way out of the condition of anti-autonomy. But how can that be compatible with modern autonomy? This is a problem if we want to reconcile Augustine's ideal of the person with the modern ideal of autonomy, and the problem gets only more difficult to solve if we look at the relation between grace and love.

3. Love, dependence, receiving

Once Augustine reaches the ideal of inner harmony, that is, after his conversion, he holds God accountable for the change. 'You have broken the chains that bound me', says Augustine repeatedly to God after his conversion (IX.1/181). What does this mean? Is it completely up to God? If not, what is the precise 'division of labour' here? And what does God do precisely? These questions are fundamental with regard to the question of autonomy. Augustine's account of freedom is most likely to contradict the (modern) ideal of the autonomous person here. If God breaks my chains, I am dependent on somebody else. This conflicts with outer autonomy, and therefore, it could be argued, with autonomy as a whole. If my inner autonomy is only possible through dependency on God, I am no longer autonomous. This argument assumes that both inner and outer autonomy are necessary conditions for autonomy as a whole. This is worth noting, since the relationship between inner and outer autonomy hasn't been touched upon yet in this context. For now I have to conclude that the ideal of autonomy is incompatible with Augustine's idea of freedom, and it is difficult to find a counter-argument here.

It could be objected that there is a way out of this problem if we try to interpret the role of God here in a Platonic way. We noted before in our discussion of Plato's ideal of the person in the *Republic* that one way to reach the good is to try to see unity in all that exists. This route is also open to the one who wants to reach the Augustinian ideal of the person. In Paul's words, it means that we can 'catch sight of God's invisible nature through his creatures' (Rom. 1:20). Augustine is aware of this route. However, overall his idea is more that of a *direct* rather than an indirect way to God, and in particular a direct *personal relationship*. Augustine's ultimate point

of reference is not 'the good' or 'the beautiful' or 'unity', but God as a person. So God is not merely a creator of good, retreating afterwards in the background, allowing for Platonic contemplation of the good. Rather God himself is the focus of Augustine's attention. He is a person. It is characteristic of persons that they can relate to (and have relationships with) other persons. Augustine's (and indeed the Christian) idea is that we are related to God. Freedom is something that emerges out of this relationship. Furthermore, it is not just any kind of relationship, but a *loving* relationship. Such a relationship contains two movements: one from person A to person B, and one from person B to person A. On the one hand, there is Augustine loving God. 'You called me: you cried aloud to me. You touched me, and I am inflamed with love of your peace' (X.27/231–2). Augustine wants to hear God, listen to him, be touched by him. This is the language of 'worldly' love, but Augustine has re-directed his attention to God, whom he longs for. On the other hand, there is God loving Augustine. Here the term 'receiving' or 'reception' is useful to discuss what is going on, on the part of Augustine. 'Let me drink you in,' he prays to God (XI.2/254–5). This is reception, which typically involves two stages. If I drink, I first put the cup to my lips and open my mouth. This corresponds to my own effort. I have to be receptive to attain the ideal. Second, the liquid flows into my body. I receive. This corresponds with receiving grace in Augustine's ideal. Now we have arrived at grace. As we see, grace is not a matter of something that happens to me without my having to make an effort. Rather, I have to attend to God, be receptive, and then grace *might* flow to me – there is no guarantee.

This ideal of the person is of course very alien to the modern ideal of the autonomous person. There is a clear suggestion of full dependency here. Augustine asks God: 'Speak to me; breath words of truth to me' (XII.10/286–7). Breathing is a metaphor that expresses dependency even more than drinking. We need air continuously. Augustine suggests that we need God continuously to live spiritually. Again, there is our own effort involved. We need to open up, take in the air, breathe. But the air itself (grace) is given to us. 'I call you to come into my soul, for by inspiring it to long for you, you prepare it to receive you' (VIII.1/311). But, as said, there is no guarantee. It is up to *God* to decide whether he gives us his grace or not. Therefore, precisely because the locus of decision is not me but God, this idea conflicts with the ideal of the autonomous person. Apparently, the Augustinian idea of grace makes autonomy impossible.

2.4. The challenge

This confrontation of modern with (selected) ancient ideals leaves us with three options. A first option would be to argue that the ancient ideals do not provide any challenge to the modern ideal of the autonomous individual, that they have nothing to do with the issue of autonomy, control, etc.; or, if they do, that the ideal involved contrasts to such an extent with our ideal of the person that the possibility of a fruitful comparison with or an influence on our ideal has to be excluded. Furthermore, one may claim that our modern ideal is superior to the 'ancient' ones; we don't need those. A second option would be to argue that the insights of the ancient ideals are so radically different from the modern one *and* so superior that we need to abandon our modern ideal completely and try to develop an alternative framework to understand what a person is. One may claim that we need to think in a radically different way about the issues of autonomy, control, freedom, etc. and adapt our ideal of a person. I will not pursue these options. Rather, I will argue that we can adapt or supplement the modern ideal of the autonomous person to accommodate the insights of the pre-modern ideals. Such a comparison is meant to result in both a better understanding of the ancient ideals and an improvement of the modern ideal. Let me explain why I regard this third option as the only option that is really open to me.

By the term 'superior' (used in describing the first two options) I mean that one of the two ideals is a better ideal to aspire to since it is better in making sense of the way most of us (moderns) understand ourselves and what we aspire to be. It is also a more coherent notion of what we (as persons) are and wish to be. In other words, it's a better philosophical account of a person and the ideal of a person. This already suggests that it is too one-sided to argue for one of the first two options; it seems reasonable to expect that both ideals have something to offer to us.

It may be noted that the first two options are similar in this respect: the gap between modern and ancient ideals of a person is just too wide, so that either the modern (the first option) or the ancient (the second option) is preferred and argued for. The third option implies trying to reconcile both views, refusing to see them as fundamentally rival views.

Given that from my discussion in this chapter *some* meeting points between 'modern' and 'ancient' emerged, the last option is the only one I will argue for. I hope to further rule out the other options in the

course of my book. This means that after having established the tension between ancient and modern ideals of the person I need to succeed in resolving this tension by constructing a plausible and acceptable ideal of a person on the basis of the existing dominant modern ideal but leavened with the best insights of the ancient one. This asks for two lines of argument. The first starts from the dominant modern ideal, but detects a fundamental problem with it; by this, I establish that there is a problem and that we ought to do something about it (this is the content of the next chapter). The second starts from the ancient descriptions of phenomena and theoretical ways of self-understanding: how can we understand from these (a) their ideals and (b) our ideal; and, therefore, (c) what are the differences precisely? (This line of argument has been started in this chapter and will be continued in Chapter 4.)

Then the two lines of argument meet: are there elements/aspects that result from my study of the ancient ideals (and my comparison with the modern ideal) that may be precisely those elements/aspects needed to augment the modern ideal to provide a fuller and rich account of what a person should be? (This is the aim of Chapter 4.)

In terms of what I have called the 'dialectical' structure of Part I this development of my argument (and of the two lines of argument that meet) can be seen as an attempt to construct a synthesis by showing that the antithesis can provide the elements that the thesis lacks; showing, in other words, not only that thesis and antithesis *can* be reconciled, that such a reconciliation is of great help to the modern ideal, but also that such a reconciliation is *necessary*, namely to solve a specific problem with the modern ideal of autonomy. In the next chapter, I shall show this problem and argue for the need to extend the modern ideal.

3
Problems with the Modern Ideal: the Need for Extension

Introduction

In the first chapter, I sketched the dominant modern ideal of autonomy. But to achieve an attractive picture (and no doubt it *does* attract many of us) I had to leave out many critical questions. I also had to be very selective in my interpretation of the primary sources. Now, I want to show that the modern account of the ideal of autonomy given so far is incomplete. There is a 'gap' that opens up when we persist in asking the question: 'If I want to be autonomous, what do I want?'. The following exposition of this gap is an essential part of the development of my argument in Part I towards a synthesis; it shows that there is a major problem with the thesis which requires an extension[17] to solve this problem.

3.1. Frankfurt and the problem of infinite regress

In Chapter 1, I presented Frankfurt's ideal of the person based on his hierarchical model of the person's volitional structure. A major problem in Frankfurt's account, however, is his insistence that, on the one hand, 'there is no theoretical limit to the length of the series of desires of higher and higher orders' (Frankfurt 1982 (1971): 91), and, on the other hand, his saying that it is possible to terminate such a series of desires 'when a person identifies himself *decisively* with one of his first-order desires' (Frankfurt 1982 (1971): 91). The rationale for the first claim is that it is always possible to form a desire of a higher order; therefore there is no theoretical limit. This implies that there can be no such thing as a 'highest order' desire. The second claim seems informed by the wish on the part of

Frankfurt to still have a limit of some sort. Apparently, Frankfurt considers the termination of an infinite series of desires as essential to the achievement of autonomy, and I agree with him on this. How is it possible to direct yourself at all if there is no limit to the formation of higher-order desires? For example, I want ice-cream. If I ask whether this is the will I want, and whether the will to question whether this is the will I want, is the will I want, and whether the will to question whether the will to question whether this is the will I want, is the will I want, and so on, I soon end up is a state that excludes the possibility of governing myself, that excludes the possibility of autonomy. Consider Frankfurt's analysis of the condition of the person who suffers from an unresolved conflict among his second-order desires:

> For it either tends to paralyse his will and to keep him from acting at all, or it tends to remove him from his will so that his will operates without his participation. In both cases he becomes, like the unwilling addict though in a different way, a helpless bystander to the forces that move him. (Frankfurt 1982 (1971): 91)

Frankfurt suggests that a person who suffers from an endless regress of desires experiences this unresolved conflict of desires too. It prevents his autonomy by making him 'a helpless bystander' who no longer participates in the operation of his will. In such a condition self-rule is impossible. Moreover, if 'it prevents him from identifying himself in a sufficiently decisive way with *any* of his conflicting desires', it also 'destroys him as a person' (Frankfurt 1982 (1971): 91). According to Frankfurt, 'the tendency to generate such a series of acts of forming desires [...] leads toward the destruction of the person' (Frankfurt 1982 (1971). Thus, we need a limit.

But is Frankfurt's concept of decisive identification a good candidate for such a limit? The way Frankfurt defines this concept is strikingly similar to the definition of what it is to have a higher-order desire. He writes about decisive identification:

> When a person identifies himself *decisively* with one of his first-order desires, this commitment 'resounds' throughout the potentially endless array of higher orders. [....] The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked. (Frankfurt 1982 (1971): 91–2)

This is very similar to Frankfurt's description of the person who 'identifies himself […], through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other' (Frankfurt 1982 (1971): 88). Frankfurt does not think that this prevents the formation of higher-order desires, so why would 'decisive identification' lead to the 'dissolution of the pointedness of all questions concerning higher orders of desire', as Frankfurt claims (Frankfurt 1982 (1971): 92)? What else is a higher-order desire than the decisive identification with a lower-order desire? It is difficult to see what 'decisive identification' would amount to else than another higher-order desire. Frankfurt does not provide a convincing argument for a distinction between these two concepts.[18] Furthermore, the concept of 'higher-order desires' itself is controversial. For example, in his introduction to *Free Will*, Watson[19] observes that 'higher-order desires are simply further desires' (Watson 1982: 7). But even if we accept 'higher-order desires' and the hierarchical model of desires as meaningful, it remains unclear how such a higher-order desire can be a limit to the formation of higher-order desires at all. Since the possibility of a highest-order desire has been ruled out by the first argument, Frankfurt's concept of decisive identification cannot be accepted as a limit. If we want to retain Frankfurt's hierarchical model of desires, this conclusion leaves us with two possibilities. The first is that there is no limit. In that case we are faced with a serious problem for the ideal of autonomy, since without such a limit, it is hard to see how autonomy is possible at all. An infinite number of higher-order desires makes self-rule impossible, as I argued above. This is also Frankfurt's motivation for creating the concept of decisive identification: he wants to uphold the ideal of autonomy. However, this concept has been shown to fail in its role as an ultimate limit to the hierarchy of desires. The second possibility is that there is indeed a limit, but since the concept of decisive identification could not provide us with such a limit, it has to be provided in some other way. Since in this book I am concerned with the conditions necessary for the ideal of autonomy to be viable and intelligible, I will search for other candidates to play the much-needed role of limit.[20]

I will engage with Frankfurt's account of autonomy at greater length and in more detail in Part II (Chapter 6). At present we can conclude that Frankfurt's account of autonomy as discussed so far needs supplementation with something else which provides a limit to the hierarchy of desires.

3.2. Taylor

Taylor's concept of 'strong evaluation' can help us to start developing an account of what this 'something else' could be. What do I need in order to be able to evaluate my life and myself? Taylor's answer is that this can only be done with reference to certain values or ideals. As an autonomous person I want those values and ideals to be my own. But what is meant by 'my own'? If I say that I question myself on the basis of something that is 'my own', this seems to be circular. The object of my evaluation (myself) cannot be the same as the basis on which to make this evaluation (what is my own). I propose that the only adequate way to understand this 'my own' is to say that surely, in a first sense, they are 'my own', since *I* am making strong evaluations, judging, etc. But, if we reflect further on what we base ourselves on in making these evaluations and judgements, the source of these values and ideals, we arrive at the view that, in a second sense, they are not my own. For example, if I decide that I should use my car less often, then I make this judgement on the basis of *my* value judgement that it is good to care for the environment. However, this value is also part of the society and culture I live in, and the idea of not using a car when I have one available may have not even come to mind if I lived 50 years ago. In general, our values appear to be 'given' to us.

Taylor recognises this. He criticises Nietzsche's view that values are (or should be) our own creations. (I will not develop this criticism further.) Furthermore, he argues against Sartre's idea of radical choice. Roughly defined, this is the idea that we can choose ourselves. We can choose what we value, and we can choose what to do with our life. This is a 'radical' choice because, according to Sartre, these choices are entirely 'up to us', they are not determined by anything or anybody else than me. Sartre gives examples of 'dilemmas' that (he believes) cannot be solved in an *a priori* way; we have complete freedom of choice. Taylor argues that although we do make choices, the values on which we base these choices are not themselves chosen by us: 'Our evaluations are not chosen. On the contrary, they are articulations of our sense of what is worthy, or higher, or more integrated, or more fulfilling, and so forth' (Taylor 1976: 122). A Sartrean dilemma is a dilemma only 'because the claims themselves are not created by radical choice' (Taylor 1976: 119). Sartre's idea of radical choice would imply that we just throw ourselves in one or other direction. The one engaged in radical choice 'just throws himself one way' (Taylor 1976: 119), since there are no grounds, no reasons to choose one direction

rather than another. If we really evaluate, however, we base our choice on reasons and values. Our choice is not 'radical' in the sense Sartre means it. The idea of radical choice appears to be incoherent. Moreover, this idea is certainly unintelligible as part of an ideal of the person. If autonomy means that we make our own choices, we want to make these choices on the basis of what we hold important, valuable, worthy, etc., and we want to use our capacity for reason rather than 'throw' ourselves in one or another direction, making a radical choice. If what I hold important and valuable is itself entirely up to me, it seems I lack a ground on which to base my choice. I will say more on Sartre in the next section and in Part II (Chapter 5).

So far, I have shown that there is a problem of infinite regress, that something is needed to stop it, and that values and ideals might be able to fulfil this role. However, I have started to argue that if these values and ideals are understood as being radically 'my own', they cannot be the limit needed to stop regress. If strong evaluation were to imply that I am able endlessly to generate higher-order desires, it becomes impossible. Taylor proposes values and ideals as being somehow 'given' to us. It remains to be seen whether this can be the final answer to the problem (I will argue below that it is not), but in any case it is clear that my account of the ideal of the autonomous person presented in Chapter 1 needs extension to deal with the infinite regress problem. Taylor rightly referred to the existentialist view of radical choice to illuminate this problem, since this view bears out what happens if we say that choice, including choice related to self-evaluation, is radically 'up to me'. Therefore before going on to develop an extension to the ideal of the autonomous person, it will be instructive further to discuss existentialism to reinforce my argument about the necessity for extension.

In Iris Murdoch's argument against what she calls the 'existentialist-behaviourist' view we also find a critique of the existentialist view of the person. Her critique and, as I will discuss later, her own constructive discussion of freedom, will help me in refining the ideal of autonomy.

3.3. Murdoch

Murdoch argues that what she calls 'the existentialist-behaviourist view' is one-sided and impoverished for two reasons. The first reason is that the inner life is neglected. The stress is on the publicly observable. The second reason is that it involves 'the elimination of the substantial self' and an 'emphasis on the solitary omnipotent will' (Murdoch

1970a: 9). Indeed, many contemporary theories of freedom reduce human being to the movement of a choosing will. Murdoch mentions Sartrean existentialism, but also Hampshire's and Hare's writings. According to Murdoch, the result is that, on the one hand, action is over-stressed. Action is treated as not only a necessary condition for a decision to be a decision; more, it is treated as if it *were* the decision. The public, observable action is important, not the private thoughts of the person that precede the action. On the other hand, we get a very strange image of the moral agent. Murdoch argues that he is pictured as 'an isolated principle of will, or burrowing pinpoint of consciousness, inside, or beside a lump of being which has been handed over to other disciplines, such as psychology or sociology' (Murdoch 1970b: 48). Since there is no rich conception of a self any more (Murdoch in fact does not explain in detail what that would amount to), what remains is on the one hand the will, and on the other hand the psyche as an object of science. Both are isolated from each other. And where is the agent? 'The agent, thin as a needle, appears in the quick flash of the choosing will' (Murdoch 1970b: 53). We see that all our freedom, all our possibilities as humans, in fact, our very humanity, is injected into this abstract idea of the choosing 'will', which makes the rest of our-selves indeed a 'lump of being', a de-humanised object, left over to determinism.

I agree with Murdoch that this image of man is 'alien and implausible' (Murdoch 1970a: 9). However, the point I want to make here is that it is especially implausible if such an image is to be part of the ideal of the autonomous person. If we wish to be autonomous persons, which indeed includes the wish to be able to choose freely or to have a free will, do we really wish to *be* a freely choosing will? In other words, is autonomy something which only relates to our *will*, or does it rather engage our self as a whole? The way I described the ideal of autonomy up till now (using Berlin, Feinberg, Christman, Frankfurt and Taylor) tends to suggest the latter. Autonomy appears to involve not just our will as such, but in fact a whole range of things which are part of, or closely related to, our self: our aspirations, dreams, plans, values, ideals, conception of ourselves, reasons, attachments, cares, etc. Nobody (to my knowledge) claims that these things are not important, but the point is that Sartrean existentialism and behaviourism hold that they are *not relevant to my choosing*, and hence to my autonomy. My discussion of problems with this view so far, however, suggests that the modern ideal of the autonomous person requires a different view of what it is to choose, decide, and judge for oneself.

Murdoch attacks in particular the existentialist idea of freedom. She argues that such a picture of freedom is simply not realistic. In our experience, moral choice is not like that. She asks: 'If we are so strangely separate from the world at moments of choice, are we really choosing at all, are we right indeed to identify *ourselves* with this giddy empty will?' (Murdoch 1970a: 36). However, not only in existentialism, but also in Hare's and Hampshire's theories of freedom,[21] the moral agent's freedom is the freedom to withdraw, survey the facts, and choose again. The question is now, according to Murdoch, 'whether the idea of the proud, naked will directed towards right action is a realistic and sufficient formula' (Murdoch 1999 (1970c): 368). After the Enlightenment, Romanticism, and Liberalism, 'we have been left with far too shallow and flimsy an idea of human personality' (Murdoch 1999 (1961): 287). The joining of a materialistic behaviourism with a view of the individual as a solitary will has resulted in an image of rational man, totally responsible for his actions, nothing transcending him, while his inner life is resolved into his acts and choices. Murdoch asks: 'What have we lost here? And what have we perhaps never had? We have suffered a general loss of concepts, the loss of a moral and political vocabulary' (Murdoch 1999 (1961): 290).

So at stake here is a critique of an idea of freedom which – whether true or false with regard to the nature of freedom – turns out to be unsuitable as an ideal of the autonomous person, if any ideal at all. Firstly, the Nietzschean or Sartrean 'ideal' of respectively self-creation and radical freedom is logically incoherent, as I suggested on the basis of Taylor's and Murdoch's critique. Secondly, it is inconsistent with the ideal of the autonomous person articulated so far.

But if the existentialist ideal of autonomy or the unextended Frankfurtian ideal of autonomy is implausible and highly problematic, the question is what extension we need. Taylor's view suggests that 'ideals' and 'values' will do as grounds for engaging in strong evaluation and as limits to the infinite regress of (higher-order) desires. But does this mean that we have uncritically to use ideals and values that are dominant in our culture and society? The modern wish to be an autonomous person, it seems, includes the wish to be able to question these ideals and values. I wish to be free to decide whether or not I endorse, identify with, and integrate in my life the ideals and values that crystallised in the course of the history of my culture and society. It seems to me that this wish is typically modern and that therefore any modern ideal of the autonomous person should account for this wish. But how am I to decide whether or not to identify with a certain

ideal or value? And how do I decide whether or not a certain ideal or value is 'mine' if it is this 'mine' that is in question (in strong and radical evaluation)? How do I decide between conflicting ideals or values? Are certain ideals or values more important than others? How can I order them in a hierarchy? I see only one solution to these problems: accepting that the ideal of the autonomous person requires a notion of an 'ultimate' desire referring to an 'ultimate' value or ideal which can not find its source in the person if it is to be coherent *as an ideal*. To add more support to this claim, I will now extend the discussion by looking at Susan Wolf's essay 'Sanity and the Metaphysics of Responsibility' (1988).

3.4. Wolf

Given what I have said about the ideal of autonomy in Chapter 1 and my discussion of problems so far, it is clear that there is a key question the person who wishes to be autonomous has to answer: What is mine, and what is not mine? What is it that *I* really want, what is the will *I* really want to have, what are *my* desires, what are *my* values, what is *my* ideal, what or who am *I* anyway? If I wish to 'rule myself and let nobody rule I' (to use Feinberg's definition again), the key question is who or what this 'myself' and 'I' is. A complete account of the ideal of the autonomous person, therefore, needs to deal with this question of identity. In Wolf's words, we need to be able 'to distinguish cases in which desires are determined by forces foreign to oneself from desires which are determined *by* one's self', and therefore we need to know ourselves, our identity. An obvious way to deal with this problem is to assume that there is something like a 'real' self; Wolf calls it a 'Deep Self' (Wolf 1989 (1988): 140). This would be a solution to the infinite regress problem and the problem of evaluating ideals and values. If there is something like a 'Deep Self' we could find out the will we really want to have, and evaluate the ideals and values that are 'given' in the society and culture we live in. But there is a serious problem with this 'Deep Self' view: is there really such a 'Deep Self', and is it really the 'ultimate' point of reference by which we are able to decide what we really want and who we really want to be? How 'fixed' really is our identity such that it is able to take on such a role?

Wolf approaches this problem from the perspective of questions concerning responsibility. According to her, 'there remains the question, who, or what, is responsible for this deeper self?' (Wolf 1989 (1988): 141). Frankfurt's 'solution' (decisive identification) and Taylor's

'solution' (the possibility to evaluate our selves who make the first evaluation – we may call it a kind of second-order evaluation) do not solve the problem, since they only create 'endless levels of depth' (Wolf 1989 (1988): 141). Wolf's main argument is that 'no matter how many levels of self we posit, there will still, in any individual case, be a last level – a deepest self about whom the question, "What governs it?" will arise as problematic as ever' (Wolf 1989 (1988): 142). Even if my actions are governed by my desires, and my desires by my deepest self, 'my deepest self will still be governed by something that must, logically, be external to myself altogether. […] I am not in control of my deepest self' (Wolf 1989 (1988): 142).

This is an important conclusion. Whatever name we want to give to what Wolf calls 'my deepest self', there is a point when I am not in control any more. This being so, however, we still retain our responsibility according to Wolf. She argues that 'although we may not be *metaphysically* responsible for ourselves – for, after all, we did not create ourselves from nothing – we are *morally* responsible for ourselves, for we are able to understand and appreciate right and wrong, and to change our characters and our actions accordingly' (Wolf 1989 (1988): 147). This capacity Wolf calls sanity. For the question of autonomy this implies that autonomy does not include the ability to create ourselves but *does* include the ability to revise, to correct ourselves. In other words, the ideal of autonomy means (among other things) that 'we *take* responsibility for the selves that we are but did not ultimately create' (Wolf 1989 (1988): 148), that we try to correct or change ourselves. But when do we have to correct ourselves, change ourselves? What is the criterion? If we have 'the *ability* to cognitively and normatively understand and appreciate the world for what it is' (Wolf 1989 (1988): 150; Wolf's italics), which norms or standards do we use? Here Wolf too touches a limit. Neither the source of our 'deep selves' nor the source of our norms is made explicit by her. For the ideal of autonomy to be coherent it needs to include such an explicit account of 'the source'.

Although Wolf doesn't say anything about the source of our deep selves as such, there is in her essay an indication about the relationship between this 'source' and the 'deep self'. She argues that we are tempted to suppose that, following the control of our actions by desires and the control of our desires by our higher-order desires and eventually our deep selves, 'we must have yet another kind of control to assure us that even our deepest selves are somehow up to us. But not all the things necessary for freedom and responsibility must be types of

power and control. We may need simply to *be* a certain way, even though it is not within our power to determine whether we are that way or not' (Wolf 1989 (1988): 144). How can we understand this? Wolf goes on to explain her concept of sanity already alluded to above, 'the ability to cognitively and normatively recognise and appreciate the world for what it is'. But why call this 'sanity', and not 'morality'[22] for example? Surely madness and morality are not necessarily opposed, as my argument in regard to Plato's view of madness shows (Section 2.2. and Section 4.1.). To use 'sanity' as the key concept seems to assume a perhaps too narrow meaning of madness.

But apart from this terminological point, there is the need for more explanation. If my purpose is to make sense of the ideal of the autonomous person (my term, not Wolf's), and therefore of the influential 'deep self view', I need an account of the 'recognition' and 'appreciation' Wolf mentions. In particular, the question about the relation between autonomy and morality needs discussion. Again, Wolf's account touches a limit. I think she offers the right critique of what she calls 'the deep self view'; but her concept of sanity is not able to fill the gap shown in the 'deep self' account. It is likely that it was not intended to have this function, and there is a difference between the issue of autonomy (my point), the issue of responsibility and sanity (Wolf's focus), and the issue of morality (related to both). But this does not render Wolf's concept of sanity less problematic. If sanity 'involves the ability to know the difference between right and wrong,' (Wolf 1989 (1988): 145) we want to know how this moral ability is possible at all, indeed how this moral consciousness (and not sanity) is possible. More precisely, we want to know more about the source, or the kind of source, that is necessary for us to get an idea of right and wrong, the kind of 'thing' that ultimately closes the chain of higher-order desires, the 'thing' that I cannot control.

It is certainly not sufficient to refer to the creation of our 'deep self' by upbringing, social circumstances, and other social/cultural elements. After all these elements (values etc.) are the *object* of our 'strong evaluations' (to use Taylor's term), and not the source that makes strong evaluation possible at all. Furthermore, since Wolf later changed the name of her view to 'the Reason View' (in her *Freedom within Reason*) it is likely that by choosing 'sanity' as her main concept Wolf wanted to stress reason. But then we want to know which role reason has in the picture. To say that 'we are able to change the things that we find there is reason to change' (Wolf 1989 (1988): 145) is not sufficient as an account of that which may be able to provide a solid ground to

stop the regress of desires and judgements from becoming groundless, of infinite depth. There may or may not be such a ground, but if the ideal of autonomy and responsibility presented in her essay is to be coherent, she needs to assume that there is one and explain its precise role; otherwise we remain stuck at the 'endless levels of depth'. I think Wolf is right in suggesting that power and control may be not by themselves suitable to describe what's going on, but what we really want then is a possible alternative or supplementary way of thinking about the source or ground and in particular its relationship to the other elements of the model – desires, judgements, evaluations, 'deep self' – elements which, in the ideal at least, *are* standing in mutual relationships of power and control. What would such an alternative or supplementary way of thinking about the ground look like? To answer this question in some way is part of the challenge of Part I.

3.5. Feinberg

Whether we choose to call it an 'ultimate point of reference' or the 'source' or 'ground' of the 'Deep Self', it is clear that it is needed if we want to hold a coherent and complete ideal of the autonomous person. To firmly establish this need, and to add to the terms helpful in talking about this need, I look now at yet another way of putting the problem.

Feinberg's discussion of autonomy shows two major problems with the ideal of autonomy articulated in Chapter 1. Feinberg, like Wolf and Murdoch, is highly critical of an approach which does not include (1) a rich conception of the self and (2) reference to 'normative standards for determining the relative worth of conflicting wants and interests' (Feinberg 1973: 167). Feinberg, too, supports supplementation of the view of autonomy embodied in the ideal of autonomy I articulated in Chapter 1, which I shall henceforth call the 'standard model'. Let's reconsider both critiques (1 and 2) in Feinberg's analysis. He argues that the concept of autonomy presupposes an adequate conception of the self; one that is narrow enough but also wide enough: 'If we strip our conception of the governing self of all its standards and values, leaving only a bare impersonal Reason imprisoned in its own royal palace, the notion of autonomy becomes empty and incoherent. […] The human subject of freedom, then, must have some substance, some normative flesh and blood' (Feinberg 1973: 160). For my purposes, it is not necessary to discuss what the self of a person actually is – what a person is. My interest here lies in the *ideal*, in particular the ideal of a person, and for that reason it is interesting that Feinberg mentions

'*normative* flesh and blood'. Why normative? I propose to understand this in terms of the claim I considered before. If the ideal of autonomy is to be coherent, we need to link the independence and self-rule of the ideal person with the norms or standards by which this person can exercise his capacity for independent decision. Neither 'outer' nor 'inner' autonomy is possible without these norms.

I argued before, however, that it is not sufficient merely to refer to 'ideals', 'values', or 'norms'. Where do these norms come from? What does it mean to say that this norm is *my* norm? With the help of Feinberg we can further clarify this problem and add support to my claim that the solution is an 'ultimate point of reference'. In particular, I want to strengthen my argument that the ideal of autonomy includes the wish to deal critically with the norms we find as 'given' in our society and culture.

Firstly, it has to be recognised that ideals, values, and norms are not our own creation, that there *is* a sense in which they are 'given'. This is, as such, not a violation of the definition of autonomy. It is not because the ideals, values, and norms of the society and culture I live in are not created by me, that I am prevented from ruling myself (no inner autonomy) and that I am ruled by others (no outer autonomy). I already argued that although in one sense they are not mine (I did not create them), they can be mine if I make them mine, if I identify with them. I have the ability to critically examine them, to engage in strong evaluation, to form higher-order desires. However, I also argued that to be able to do this, I need some 'ground' or 'ultimate point of reference'. This makes it possible to achieve inner and outer autonomy. I achieve inner autonomy by putting an end to the regress of desires and in this way reach inner order and harmony. I know what I want. But why, and how, does this further outer autonomy? My answer is that a 'ground' or 'ultimate point of reference' allows us to achieve outer autonomy by allowing us to achieve relative independence from 'opinion', that is, from the beliefs, ideals, values, and norms in our society and culture (the dominant ones as well as the others). I will now clarify this point with the help of Feinberg.

I use the term 'relative' independence, since it is nonsense to say that we can be absolutely independent from our culture and our time. Consider the existentialist notion of 'authenticity' versus Feinberg's notion of 'authenticity'. Why talk here about 'authenticity'? I have argued already that the ideal of the autonomous person includes the wish to question my identity. This wish can now be understood as following from the wish to be myself. The wish to question myself is

informed by the wish to be myself, to be 'authentic'. As I noted already in relation to the identity problem, this prompts the question whether there is a 'real' or 'deep' self I can use as a point of reference. But whether or not it exists, the modern ideal of autonomy seems to assume it exists if it includes the wish to 'be myself'. If I wish to rule 'me', it seems, I first have to want to search for this 'me' that is to be ruled by 'me'. What can such a search mean in relation to the question of outer autonomy? What is this 'me' as opposed to 'society'? What does it mean to say 'I wish to rule myself and not be ruled by someone else'? The Sartrean concept of authenticity suggests that we can create ourselves *ex nihilo*. But how can we create ourselves 'out of the blue'? I agree with Feinberg that 'there can be no magical *ex nihilo* creation' (Feinberg 1973: 166); we need some 'given' elements. It is more plausible to say that we are to some extent dependent on our culture and our time, and that we need this dependence to be able to 'create' ourselves, to 'construct' our identity. The values and norms we find as 'given' are then the elements we need in this construction. Feinberg seems to recognise this when he argues that if we wish to be autonomous we should not demand total independence from the culture and time we live in. He argues that

> we must not demand total transcendence of the culture of one's time and place, for the autonomous Reason even of the authentic man will be at the service of some interests and ways of perceiving the world that are simply 'given' him by the *Zeitgeist* and his own special circumstances. […] We may all be, in some respects, irrevocably the 'products of our culture,' but that is no reason why the self that is such a product cannot be free to govern the self it is. (Feinberg 1973: 162)

However, Feinberg's argument is slightly different from the one I wish to advance. According to him, it seems, the self is 'given' and then it's up to me to govern myself. I maintain that such a self-government includes the possibility to question the 'given' self and (if necessary) change it. Feinberg puts less stress on radical self-evaluation in his concept of autonomy. Although he writes that 'my modest point is simply that a person must already possess at least a tentative character before he can hope to *choose* a new one' (Feinberg 1973: 166), he suggests that we can do with what is given to us. He argues that 'rational reflection […] presupposes some relatively settled convictions to reason from and with' (Feinberg 1973: 166) and I think it is important to stress this 'relatively settled'. I have argued that we can question and evaluate

the person we are, and that for this strong evaluation we can't merely rely on the values and norms that are given in the society we live in. The fact that there are dominant values and norms reveals that there are values and norms that are not dominant. Different values and norms can conflict. Which shall I refer to in questioning my 'relatively settled' convictions? Even if there were an entirely homogenous value pattern, the wish to evaluate it belongs to the modern ideal of autonomy. (Partly because of this ideal, we actually don't have an entirely homogenous culture in modern societies.) I have suggested that to engage in this evaluation (of ourselves and of the norms and values of our society) we need an 'ultimate' point of reference.

The need for such a point of reference to be able to reach outer autonomy in particular can be illustrated with Feinberg's example of the fact that many 'youths, eager to be authentic, still keep cultishly attuned to one another' (Feinberg 1973: 165). If this observation is right, then the question (that Feinberg doesn't answer) is: Why does this happen? Based on my previous argument, I suggest the following answer. If an ultimate point of reference is lacking, we are more prone to adopt the culture of (part of) the society we happen to live in, since we lack the means to engage in the evaluation of ourselves and the society we live in. Outer autonomy as the capacity to engage in this evaluation presupposes, firstly, (the acceptance of) my dependence on society and on myself, since I find elements I didn't create myself. I find certain values and norms that are the product of the history of the culture I live in and I find myself as the product of my personal history. Secondly, it presupposes (the capacity to reach) independence from society and ourselves, since I can rely on an ultimate point of reference by which to evaluate the values and norms of society and the self that are 'given' to me. With regard to my behaviour as a result of the evaluation of norms, this does not mean that as an autonomous person I am always a dissident; I can choose to conform (or not) (see also Feinberg 1973: 165).

3.6. 'Doing what you want' and the relation between freedom and autonomy

My account of the ideal of the autonomous person arrived at up till now is far removed from the ideal of 'doing what you want'. Consider John Ruskin's portrayal of the fly in his book *The Queen of the Air* (Ch.3):

> I believe we can nowhere find a better type of a perfectly free creature than in the common house fly. […] There is […] perfect independence

> [….] You cannot terrify him, nor govern him, nor persuade him, nor convince him. He has his own positive opinion on all matters; […] free in the air, free in the chamber – a black incarnation of caprice[…] – what freedom is like this? (Ruskin quoted by Feinberg 1973: 154–5)

This passage expresses very well what is wrong with the ideal of 'doing what you want' if this ideal means the wish to enjoy caprice. In the first chapter I have shown that the ideal of autonomy is very different from the wish to enjoy caprice. If we say 'I wish to do what I want', we do not really want the freedom of Ruskin's fly. We want autonomy, that is, we want the following. Firstly, we do not merely have desires, we also want to form higher-order desires about these first-order desires, something the fly is not able to do – as a person I want to be able to decide whether the will I have is the will I want (Frankfurt). Secondly, we want to be able to engage in strong evaluation (Taylor). Third, we want to reach inner order and independence from others (inner and outer autonomy).

The freedom of the fly is meant as freedom of action 'without constraints', but turns out to be a condition of unfreedom, since the fly is the prisoner of its first-order desires, which are the only desires it can have. Autonomy, on the contrary, presupposes constraints. Since I'm a person, I'm able to form higher-order desires, and to make use of this freedom and rule myself rather than be the victim of endlessly regressing higher-order desires or be ruled by other people, I have to rely on something that is not 'me'. That is my conclusion of this chapter.

3.7. Conclusion: the gap

I have pointed out the problem of infinite regress. I considered the following answers to this problem: ideals and values (Taylor), the source of the Deep Self (Wolf), and normative standards (Feinberg). All these elements are 'not me'; rather they are given to me as part of my personal and social environment. Still, they enhance rather than limit my autonomy: if I critically evaluate them, they can help me to be myself and therefore rule myself. This critical evaluation (which we could call a meta-strong evaluation or a strong evaluation of a second order), however, is only possible if we have an ultimate point of reference. This is the second part of the answer to the problem of infinite regress. I need such a point to be able to reach inner and outer autonomy. My conclusion is that we need to extend the ideal of autonomy articulated in Chapter 1 with the elements suggested in this chapter. It is not

clear, however, what an 'ultimate point of reference' is. So far, I have used this term merely as a name for something that could fill the gap identified in this chapter. In the next chapter I will attempt to fill the gap by using Plato and Augustine. My aim is to describe what I shall call an 'extended' ideal of autonomy on the basis of a reconciliation of ancient and modern ideals of the person. In other words, my aim is a synthesis. To reach this aim, I have first dealt with the thesis. Now I shall turn to the antithesis and see which elements from Plato and Augustine could solve the problem with the thesis.

4
Using Plato and Augustine to Fill the Gap

4.1. Plato

The Phaedrus *again*

What would it mean to fill the gap I have identified in what I called the 'standard model' with Plato's concept of 'the good'? Given the account of the Platonic ideal(s) of the person I presented in Chapter 2, it seems that 'ancient' and 'modern' are largely irreconcilable. At first sight, to use 'the good' as the ultimate point of reference seems not to be an ideal of *autonomy*. I do not wish to be dependent on anything like 'the good' to rule myself. But given the problems with the 'standard' model identified in the previous chapter, in particular the need to fill the gap identified, it is worth reconsidering my argument of Chapter 2.

With regard to Plato's ideal of the person in the *Phaedrus*, I already noted a similarity between the modern ideal of autonomy and the ideal of the person embodied in the story of the charioteer and the horses. The ideal itself, it seems, is that of 'inner autonomy'. It is an ideal of inner harmony and order, and an ideal of self-mastery. Both are compatible with the modern ideal. I agreed with Dilman's claim that the Platonic condition of self-mastery is one of autonomy: 'the person will have achieved *autonomy*: what he does will be what he wants to do, not what he is forced to do, and he will be wholly behind it' (Dilman 1999: 39). But the problem I noted was that, according to Plato, this self-rule (the charioteer managing the horses) is only possible if the charioteer has a *vision* of *the good*. But is this 'problem' not a possible solution to the infinite regress problem? Perhaps Plato provides something that fills the gap in Frankfurt's account as presented so far. Consider the following picture of an extended Frankfurtian model.

We may want to add the following elements to the Frankfurtian scheme. Firstly, there is a theoretical end-point to the 'endless' struggle between desires of various orders, whereby one higher-order desire (D*x*) is controlled by a desire of again one order higher (D*x* + 1). (We may suppose that in Plato's metaphor this would mean an endless hierarchy of charioteers.) Secondly, this end-point is a state of harmony of the soul, whereby the various desires are integrated and made 'my own'. Thirdly, the person can reach this state if and only if the person attends to the good, and is consequently moved by the good.

With regard to the third element, suppose for the moment that reason plays an intermediary role here. What do I mean by 'intermediary'? In his 'Introduction' to the collection *Free Will*, Watson makes the point (noted in note 15) that 'human freedom cannot be understood independently of the notion of practical reason or judgement, and that this notion is bound up with a distinction between desiring and *valuing*. Reasons for action derive from one's conception of *a good way to live*' (Watson 1982: 8). Now I propose to understand this 'good way to live' simply in terms of strong evaluation: I judge my desires with the help of reasons, and these reasons derive from the values I have. So this is what I mean by the mediating role of reason: reasons are 'between' desires and values. But there is more. Firstly, by referring to 'a good way to live' Watson suggests that my values are somehow structured in a coherent whole: my concept of the good life. Secondly, where Watson stops I continue by asking the questions (1) how do I structure my values (including how I place them in a hierarchy) and (2) where do my values come from? The answer to both questions could be that I receive my values from the society and culture I live in. But then it may be objected, as considered previously, that I still need a way of judging these values, ordering them in a hierarchy, etc. In other words, I do not just want to conform to the values of my society and culture, I want to be autonomous and decide for myself whether I make this or that value incorporated in my culture 'my own'. Now to be able to do *that*, I need an independent point of reference, a standard by which to make this strong evaluation of my strong evaluation (a strong evaluation of the second-order, so to speak). I want to evaluate my conception of the good life. And *this* condition of autonomy can be reached only if I have not just a conception of the good life, but also a 'conception', or rather a *vision*, of the 'good'. The good is the independent standard needed to judge what a 'good' life is.

There may be other ways of conceiving of an 'ultimate' point of reference, but the Platonic concept of 'the good' seems to be an excellent candidate for an extended ideal of the autonomous person.

I will now refine this ideal further, explaining (1) attention and (2) being moved by the good. I will also further develop the relationship suggested between the Platonic ideal, on the one hand, and the modern ideal of autonomy, on the other hand. This is essential in view of my efforts to reconcile them and to arrive at a synthesis.

Murdoch and Platonic vision

We have seen how Murdoch critiques the existentialist view of freedom and (therefore) of autonomy. However, she offers not only a critique, but also claims to defend a *different* view of freedom. It remains to be seen whether this is indeed a different view or not, and whether it contributes to a different view of *autonomy*. Our discussion up till now suggests the necessity of an extension or supplement, not an alternative, to what I have called the standard model. But in any case, Murdoch's discussion of freedom may tell us something about the Platonic ideal of a person, and give us the opportunity further to expand the points made above, in particular since her ideal of the person is based on Plato's.

Murdoch claims that 'freedom is not the sudden jumping of the isolated will in and out of an impersonal logical complex, it is a function of the progressive attempt to see a particular object clearly' (Murdoch 1970a: 23). What does this claim tell us about (1) the Platonic ideal and (2) the modern ideal of autonomy? Firstly, it allows us to interpret the Platonic ideal of the vision of the good as the endpoint of progressive attempts to reach this ideal during a long period of time. Murdoch lets in 'the historical individual':

> The idea of 'objective reality' is to be understood, not in relation to 'the world described by science', but in relation to the progressive life of a person. The active 'reassessing' and 'redefining' which is a main characteristic of live personality often suggests and demands a checking procedure which is a function of an individual history. (Murdoch 1970a: 26)

Note that I already hinted at something like 'individual history' when arguing that just as we are faced with a historically developed society as something that is 'given' to us, our self is similarly 'given' and historical. Note also that Murdoch's and Plato's notion of seeing

'reality' roughly corresponds with Wolf's notion of 'the *ability* to cognitively and normatively understand and appreciate the world for what it is' if we understand 'the world' as including my *personal* reality, my personal world.

Secondly, with regard to autonomy, we may infer that to reach the ideal of autonomy is not the work of an 'instant' moment, but takes time. Just as, according to Plato, to reach the idea of the good takes time (as I argued in Section 2.2.2.), the process of trying to reach autonomy has a continuous character. Frankfurt's model may be too easily understood in terms of 'decisive moments', since his concept of 'decisive identification' (see also Chapter 3) can be interpreted as requiring a single act of will: I decide here and now which desire is to be 'mine'. Within Frankfurt's model 'decisive moments' are those in which a volition of a higher order decisively forces itself on the lower-order desires to bring about order. But such moments of choice may be only part of what's going on, and certainly only part of what is ideal. 'At crucial moments of choice,' Murdoch writes, 'most of the business of choosing is already over. This does not imply that we are not free, certainly not. But it implies that the exercise of our freedom is a small piecemeal business which goes on all the time and not a grandiose leaping about unimpeded at important moments' (Murdoch 1970a: 37). Therefore, with regard to the question of how to reach autonomy, I infer from Plato and from Murdoch that self-rule too is not something that involves being focused on certain moments of explicit 'choice'; what happens between such 'choices' may be much more crucial. If autonomy involves self-evaluation and the evaluation of societal norms and standards, this is not something that is or can be done in one 'act of will'. To describe these processes leading towards autonomy we could use Murdoch's concepts. Murdoch deploys as her key concepts *attention*, *looking*, and *vision*. 'If I attend properly,' she writes, 'I will have no choices and this is the ultimate condition to be aimed at' (Murdoch 1970a: 40). It is a condition of 'necessity'. If we are free in this sense, we do not have to choose at all anymore, since we 'know' what to do, because we attend to the right thing. With regard to autonomy, this is even the *ideal* case: if attention to 'the right thing' can help us in 'knowing what to do', this means that it will be easier for us to reach a state of inner order and harmony between possible conflicting desires, reasons, and values. If I know what it is right to do, I am able to say: 'I rule me and nobody else rules I'. If, on the other hand, I do not know what is right, such a phrase would be senseless. Autonomy as self-determination seems to include *determination as*

knowing what to do as a consequence of knowing what is 'right'. But if I do not know what to do, the very possibility for determination and therefore self-determination is ruled out. Murdoch's concept of attention adequately describes one of the conditions for the ideal of the autonomous person.

Murdoch connects attention with love. She writes that 'the idea of a patient, loving regard, directed upon a person, a thing, a situation, presents the will not as an unimpeded movement, but as something very much more like "obedience"' (Murdoch 1970a: 40). 'Pure will,' Murdoch argues, 'can usually achieve little. It is small use telling oneself "Stop being in love…" What is needed is a reorientation which will provide an energy of a different kind, from a different source' (Murdoch 1970b: 55). Note the difference between this concept of reorientation and Frankfurt's concept of 'decisive identification'. Exercising my freedom, in Murdoch's view, appears to have little to do with the identification with one of my desires, but a reorientation towards something different. Instead of what we could call an 'inward' move of control, Murdoch shifts the focus to an 'outward' move of attention. And even if we would want to retain Frankfurt's concept of decisive identification, Murdoch's concept of attention suggests that in order to be able to identify decisively with a particular desire, we do not need an 'act of will' as such but vision and reorientation. Notice the metaphors of orientation and of looking: 'Falling out of love is not a jump of the will, it is the acquiring of new objects of attention and thus of new energies as a result of refocusing' (Murdoch 1970b: 56). In Murdoch's view, freedom, then, 'is not strictly the exercise of will, but rather the experience of accurate vision which, when this becomes appropriate, occasions action' (Murdoch 1970b: 67).

It may be objected that what I have called the 'outward' move, a shift of attention as opposed to 'inward' control, may be or may not be related to freedom, but in any case makes the ideal of *autonomy* problematic if we extend this ideal with Murdoch's Platonic concepts. If Murdoch's concepts are meant (by me) to express part of what it is to be autonomous, is it still the case as I said just above that 'I rule me'? If to be autonomous is to (be able to) attend to an object, is it not the case then that that object rules me? My answer[23] to this question is that – according to the Platonically extended ideal – the ultimate 'object' relevant for the question of autonomy is not just any ordinary object, but some *good*, and ultimately *the good*. With regard to autonomy, the difference between an ordinary object and the good is that the latter alone can provide a point of reference that can help me to

reach inner harmony and outer independence. Objects may *rule* me (or I may rule them) but the good alone is able to *guide* me. By attending to the good, I avoid a situation of having no reference point to guide my decisions, actions, choices. I can rule myself because I know what to do, and I know what to do since I have vision of the good. Attention to objects, by itself, is unable to end conflict between, and regress of, (higher-order) desires. Only (attention to) the good succeeds in supporting and guiding a commitment that, in Frankfurt's words, 'resounds throughout the potentially endless array of higher orders' (Frankfurt 1982 (1971): 91), making decisive self-rule possible. The Platonic assumption is that we should not attend to ordinary and imperfect objects, but to ideas lying beyond the world (of objects), and in particular to the idea of the good. Murdoch's source of inspiration for her 'new' concept of freedom is Simone Weil, and, ultimately, Plato. She sees the idea of the good as the source of light which reveals to us reality: 'Good is the magnetic centre towards which love naturally moves. […] Love is the tension between the imperfect soul and the magnetic perfection which is conceived of as lying beyond it. […] It is the energy and the passion of the soul in its search for Good' (Murdoch 1999 (1970c): 384).

Murdoch helps us to better understand Plato's ideal of the person in Book VII of the *Republic* and therefore helps us to add the following elements to the extended ideal of the person, that is, the ideal I developed so far in this chapter on the basis of the *Phaedrus*. Firstly, the ideal 'to see the Idea of the good' is reached only at the end of the journey upwards. To use Murdoch's terminology, there is a 'personal history', since, as Plato says, 'in the world of knowledge, the Idea of the good appears last of all, and is seen only with an effort' (517b). For the extended ideal of the person, this means that radical (self-)evaluation is something that takes time. Secondly, as noted already in Chapter 2, the ideal of 'vision of the good' is not contrary to rationality.

Rather, Plato argues that one who wants to act rationally must have this vision of the good. The Idea of the good is 'the power upon which he who would act rationally either in public or private life must have his eyes fixed' (517c). To use Murdoch's idiom, proper rational action presupposes attention. With regard to autonomy, vision of the good is a necessary condition of rational self-governance; that is the rational evaluation of both one's own norms and values and those of the society within which one lives. (See also the intermediary role of reason mentioned above.) Thirdly, Murdoch's idea of redirection is consonant with Plato's view that the (moral) problem is a question of

direction. As I said in Chapter 2, Plato holds that we all have the faculty of vision, but the question is whether or not we turn to the good. Murdoch understands this turn as the redirection of our attention to the good.

What does this problem of direction mean in relation to the modern ideal of autonomy? And why is a Platonic extension, with 'the good' as something we should turn to, *necessary* anyway? How is the ideal of the one who acts and decides rationally compatible with the modern one? I identified the gap in the modern ideal, but given what I have said in Chapter 2 it is not yet fully evident why we should want to consider the Platonic ideal to fill the gap.

Merging ancient and modern ideals of the person

I shall argue that it is not only desirable but also possible – and necessary – to supplement what I have called the 'standard' model of autonomy with features drawn from the Platonic ideal of a person. To start thinking about what such a supplement would look like, and to argue why it would be a coherent ideal of the person rather than a conceptually incoherent one, we may want to consider the following metaphor.

Imagine a captain on a ship who is given absolute freedom to decide where he wants to go. The captain is free to set his own aims. It is clear that without a compass or any other means of navigation, the captain is 'free' to go where he wants to go but will not be able to exercise this freedom. The movements of the ship will express caprice. The captain will enjoy the same freedom as the fly in Ruskin's story. Although the ship is not necessarily ruled by the waves – the captain can give orders to steer his ship in one or the other 'direction' and is in that sense 'in control' of his ship – it is hard to maintain that the captain is fully autonomous. What do self-rule (the rule of his ship) and self-determination mean if he doesn't have a means of navigation? Without a compass, the notion of 'direction' becomes meaningless, and therefore self-direction (here directing the ship) too. If he has a compass, on the other hand, the captain knows where he's going and can therefore not only set his own aims, but has also a good chance of reaching them. Ruling and directing the ship make sense now. But what do a compass assume? What is a compass? In its simplest form, it is a needle that points in a certain direction, in particular to the magnetic (north)pole. Therefore, a compass is only of use in virtue of its being attracted by the magnetic pole. Without such a pole, it would be useless as a navigation tool.

Similarly, I argue that as persons we can only exercise our capacity of self-rule and self-direction if we have 'a compass', that is, if we have reason. But just as a compass assumes the existence of a magnetic pole, the faculty of reason presupposes the existence of a 'magnetic pole' that attracts it. Having looked at the Platonic ideal of the person, we could give this 'magnetic pole' the name of 'the good'. If I direct myself to the good, I can exercise my capacity of autonomy and reach the condition of autonomy. If I have a sense of (normative) direction, it makes sense to set my own goals, to direct my life in the way I want. Without such a point of reference, I have no means to prevent myself from being ruled by other people's aims, norms, desires (no outer autonomy), or by my own uncontrolled desires, lacking inner harmony and order (no inner autonomy). To use the metaphor of the boat again: If the captain doesn't know where he's going, how can he rule the crew members?

I will now further clarify my argument by considering the following objection. It could be argued that we moderns can do without any 'magnetic pole', that reason is enough. To answer this objection, let me recapitulate my argument. I am not saying that we *have* to accept the existence of such a 'thing' as 'good' (or, as we will consider later on in this chapter, God), but, rather, that we have to accept the existence of it *if* we want to uphold our ideal of the autonomous person. The alternative, many 'magnetic poles' or none at all, may be an ideal according to some people but it is incompatible, I argue, with the ideal of autonomy.

Let us first consider the latter possibility: no 'magnetic pole' at all. The reason why this is incompatible with the ideal of autonomy emerges from the problem of an endless series of hierarchical volitional controls which needs a limit if the person wants to attain the ideal of inner autonomy (inner order, harmony between the desires) (see my argument in Chapter 3). For the same reason there is a problem with many 'magnetic poles'. This is not only a contradiction in terms because of the metaphor I chose (there is only one north pole); if there are many poles, many points of reference, one would need yet another point of reference that stands 'above' these poles to be able to choose between them. Furthermore, it can be argued that 'outer' autonomy too needs a 'magnetic pole', since otherwise authentic independence vanishes into caprice ('doing what I want' – the ideal of the 'fly', as I explained above) or conformism (in the sense of the unreflective and uncritical acceptance of the values, norms, and ideals of others, of my society, or of my culture). Instead, I need an idea of the good so that I

know what is good for me, and so that nobody else has to tell me what good is for me.

Why do I mention again 'the ideal of the fly' (see Section 3.5)? My point about 'outer' autonomy shows the implication of my argument for the ideal of 'doing what you want'. Using the metaphor again, imagine a captain who desperately wants to 'be autonomous', and therefore does what he wants. He doesn't need the sun, the stars, or the magnetic pole, so he claims, and he won't follow any of them. The only point of reference is my own will, he says. This may be an option. However, it is not difficult to see that this is a meaningless ideal of autonomy. Following the analogy, the ship would follow the will of the captain, but if this will is completely unimpeded and unconstrained the situation is that of caprice. It is this that we may call 'the freedom of the fly' (referring again to Ruskin's story). This may be a misunderstanding of the reality of a fly, but in any case it expresses quite well what is meant here because that's at least the image we have of a fly, going wherever it feels like going in our room, without having any aim or point of reference (that's at least how it appears to us). This may be freedom in one sense (freedom of action), and unfreedom in another (I am ruled by my first-order desires and lack inner order and harmony; caprice is the prison of the one who does what he wants without any constraints), but it is certainly not the modern ideal of autonomy.

Note that the metaphor (this ideal of autonomy) allows for ships (persons) to go in different directions. Indeed, at no point is it suggested that we should all head in the same direction, follow the same course of life. Rather, I argue that whatever course we take, whatever direction our ships choose to go (we have autonomy in the sense that we are free to choose the direction of our lives), this choice and autonomy is only possible on the basis of navigation with reference to a certain 'magnetic pole'. Using Plato, we can identify this 'magnetic pole' as the good, the Platonic name for something that helps us in answering the question, 'Which desires, norms, and values are mine?' and therefore the question, 'What should I really desire, what should I make my norm, what should I value?'

I will now again discuss madness, including the nature of Dionysian madness and certain elements in ancient Greek culture, to achieve a more complete picture of the Platonically extended ideal of the autonomous person. It is necessary to draw clear lines between this extended ideal including elements of ancient Greek culture that are compatible with, and even necessary for, autonomy, on the one hand,

and ancient elements that do *not* contribute to autonomy, on the other hand. This will mean some repetition of the content of Chapter 2; but, although I will look at the same material, I will do so in view of a different aim (reconciliation) and with the problem identified in Chapter 3 in mind (the gap and the need for a point of reference).

Madness

My argument for compatibility so far was based on Plato's metaphor of the charioteer and the horses in the *Phaedrus* and his ideal of vision of the good in Book VII of the *Republic*. It is far less clear why his view on madness in the *Phaedrus* allows for an argument for compatibility. As I have argued in Chapter 2, there is a form of Platonic madness that is perfectly compatible with the Platonic ideal of inner harmony, namely the madness of the person who recollects and is inspired by the Idea of the beautiful, the madness of the one who 'forgets earthly interests and is rapt in the divine' (249d), the madness of 'him, who, when he sees the beauty of earth, is transported with the recollection of the true beauty; he would like to fly away, but he cannot' (249d). Similarly, I believe, the Platonically extended ideal of autonomy developed so far is compatible with the madness of the one who is inspired by the Idea of the *good*. If madness inspired by the good allows the charioteer to manage his horses, if this madness makes autonomy possible, then it is not contrary to the ideal of autonomy if, by analogy with the beautiful, I am a mad lover of the good who recognises the good on earth and is for this reason regarded by others as 'a downright madman' (251a).

With regard to the gap identified for the modern ideal, this means that the recognition of an ultimate point of reference (by which I can achieve inner order and harmony as well as outer independence) may co-exist with a certain form of madness. If the ultimate point of reference is to play the role I suggested, it is necessary that the person truly recognises this point (here: the good) *as* ultimate, as having the authority to function as the ultimate reference point he can use to self-direct his desires, aims, etc. and to reach inner harmony. Madness, rapture, and inspiration, then, belong to this sense of recognition (of the authority of the good), perhaps a mixture of a sense of awe and a sense of attraction. This does not mean that the autonomous person needs to be constantly 'mad' in this sense, and madness comes in degrees. It is possible that too much Platonic madness is counterproductive (the 'product' being autonomy), namely if the person loses himself (I will return to the issue of mystical union below). But if Platonic madness is understood as the feelings and condition that accompany the true

recognition of the ultimate reference point necessary for it to perform its function, it enhances autonomy.

By contrast, Dionysian madness, namely the condition of being taken over by a god or a daemon, is incompatible with the ideal of autonomy for the reasons I discussed in Chapter 2. If I am possessed by Dionysus, I am ruled by Dionysus (no outer autonomy) and I don't rule myself anymore. This may have been an ideal of the person in ancient Greece; it is neither a modern ideal of the person nor an ideal of autonomy. We generally do not wish to be mad in this sense and neither we nor the ancient Greeks would consider a person to be autonomous if he were in a condition of Dionysian madness.

But are Platonic madness (as we may call the madness inspired by the good) and Dionysian madness not similar in their relation to a 'oneness'? I would like to recall and elaborate the argument I made in Chapter 2 in view of the arguments of Chapter 3 identifying the gap in the standard model and my aim to reconcile ancient and modern ideals of the person (to fill the gap). Consider Nietzsche's concept of a 'mysterious primal Oneness' and the idea of mystical unity he relates to Dionysian madness. Does the Platonic lover of the good not desire the oneness of the good? The difference between Dionysian madness, on the one hand, and Platonic madness and the extended ideal of the autonomous person, on the other hand, is that the latter includes *reference* to a Oneness, namely the good, but does not include a mystical unity with the good. To be in a condition of mystical unity would mean that the person loses his autonomy since he loses the 'I' (his identity) that is supposed to rule himself. If there are no longer borders between 'me' and 'not-me', there is neither identity nor autonomy. In the Platonically extended ideal of autonomy, by contrast, the reference to the good is made by the person as an independent, 'outer'-autonomous agent. There is a 'going out of oneself' only in the sense of a contemplation of 'the one', and in the sense of accepting something else as the source of the values and reasons I decide upon. It is love, perhaps, but then a love *directed towards* the one etc., not a love that unites the person with the object of his love. It is *me* who, with the help of my vision of 'the good', identifies with a particular desire, is the subject of strong evaluation, has second-order volitions, etc. The mystical aspect of Dionysian madness, on the contrary, includes the idea of a person being *not* himself. In the oneness of mystical unity the person loses his autonomy completely. This may be an ideal or not, but if it is, it is in any case not the ideal of the autonomous person (extended or not).

However, there are other elements of ancient Greek culture that may well be compatible with the extended modern ideal. For example, the idea that there is a pre-existing order in the universe, as Kitto believes this to be part of ancient Greek culture, is compatible with modern autonomy. A pre-existing order excludes the ideal of 'doing what you want', of absolute freedom, but it does not exclude autonomy. There will be always a limit to 'doing what you want', whether because of the gods and their laws, as the ancient Greeks believed, or because of the complexity of *modern* life with its man-made order (including laws). Furthermore, we are as much subject to natural laws as ancient Greeks were. But if we understand autonomy in terms of the extended model developed so far, there is a way of thinking about this problem both satisfying the demands of Plato's ideal and the demands of the extended modern ideal of autonomy.

The case for Plato's ideal is the easiest one. Plato's ideal itself presupposes what Murdoch calls 'reality', which may be interpreted as the order of the universe, including the gods and their laws. When I have vision of 'the good', I am able to *see* reality, and the danger of *hybris* is minimised. If I do what is 'good', this includes not acting against the order of 'life' or 'the universe'. If my compass (reason or love of the good) is reliable, I will be able to find my way through the order without 'disturbing it'. The idea of the good is my guide, my magnetic centre.

The modern ideal seems, at first, to be less compatible with the idea that I can't control everything because there is a pre-existing order. It could be objected that the belief in God (see the next section) or gods cannot be assumed any more today. This is an acceptable point. However, this is not an argument for the claim that we shouldn't assume any pre-existing order at all any more. The reason why we should still assume such an order is that it is necessary if we want to retain a coherent ideal of the autonomous person. Firstly, it is necessary to accept the belief that there are laws in this universe which control my physical movements as well as all life around me. This is the basic belief modern science holds (as it tries to discover the laws of the universe), and if I am an ideal autonomous person, I will still be dependent on these laws. Secondly, I also have to accept my dependence – as an autonomous person – on an order of which the values and reasons I have and decide upon are part, an order which, by virtue of being an order, orders these values and reasons. If I attend to the good, or to reason, or to whatever name I want to give to the 'magnetic centre', I will

'know what to do' because I have a vision of this order, I *see* which value is to be preferred over another, which reason is to be preferred over another, and therefore I know what I want my desire to be, I am able to form second-order volitions.

To conclude, apart from the aspect of Dionysian madness, Plato's ideal of the person is in many ways reconcilable with the modern ideal of autonomy. It was not my aim to argue that the Platonic ideal or the Platonically extended modern ideal is an ideal (let alone the *best* ideal) we should all aspire to. Rather, using Platonic materials I have shown that 'ancient' and 'modern' can at least partly be reconciled to answer a specific problem within the modern ideal of the person interpreted according to the standard model. I will now try to show the same using Augustine. Again, my discussion will involve a dialectic between 'ancient' and 'modern', in particular between Augustine's view of freedom and his ideal of the person, on the one hand, and the modern ideal of the autonomous person, on the other hand. The aim of this dialectic is to achieve a synthesis, a reconciliation of ancient and modern ideals of the person, by using aspects from the ancient ideal(s) (antithesis) to fill a gap in the modern ideal (thesis).

4.2. Augustine

I argued in Chapter 2 that there are certain aspects in Augustine's ideal that are compatible with the modern ideal. In particular, I used Dilman's interpretation of Augustine to show that there is in *On Free Choice of the Will* an ideal of autonomy, namely the wish to be the author of your intention and decision to act and the wish to be wholehearted, not inwardly divided. However, I also argued that there is a major problem with Augustine's reference to 'the law we know to be eternal' (I, 18–19) and to God: How can I be called autonomous if I am dependent on such a divine law and/or on God?

Having articulated a plausible Platonic extension of the modern ideal of autonomy, however, I will now try to construct an Augustinian extension and show that although there remain major problems, we might want to add such an Augustinian extension to the modern ideal of autonomy to solve its infinite regress problem. I do this by comparing the Platonic and the Augustinian ideal and discussing the problems related to a possible reconciliation between specific Augustinian features such as 'the freedom to choose evil' and 'grace', on the one hand, and the modern ideal of autonomy, on the other hand.

Plato and Augustine

As I have shown already in Chapter 2, Augustine's ideal of the person has much in common with the Platonic ideal. Self-direction towards 'the one' is held to be necessary to reach inner order and harmony. And reason is not the ultimate point of reference itself, but rather is our highest and most excellent[24] faculty which allows us, in Platonic terms, to get from 'many' to 'the one'. To this extent, my argument for a Platonic extension supports an Augustinian extension as well. The hierarchy of desires is not endless, but is limited by an ultimate point of reference, a 'magnetic pole', something that is not 'me' but helps me reach inner order and outer independence. Consider the following Platonic interpretation of the *Confessions* as Augustine's account of how we can reach autonomy.

To limit the hierarchy of desires, an 'act of will' (VIII.8/171), as Augustine first proposes, won't do. To say that by an act of will I can 'decisively identify' with a desire falsely suggests that a single decisive determination will do. If we read the *Confessions*, we see that instead there is a long process involved, a process of re-directing one's attention to the good. This takes place over an extended period of time. (Note that I leave grace out of the picture for the moment; I am looking at what for now could be a Platonic process.) Here it is useful to remind ourselves of Iris Murdoch's concept of a personal history. As I said earlier in this chapter, she argued that 'freedom is not the sudden jumping of the isolated will in and out of an impersonal logical complex, it is a function of the progressive attempt to see a particular object clearly' (Murdoch 1970a: 23). Augustine's achievement of autonomy too, it could be argued, is not a sudden act of will but a continuous attempt to see the good. His belief in God requires him to identify this ultimate point of reference with God, or rather, to identify God as the source of 'the good'. This 'shifts' the focus point of (Platonic-Murdochian) attention from 'good' to 'God'. According to Murdoch, prayer is basically 'attention to God, which is a form of love' (Murdoch 1970b: 55). This attention to God is, in Augustine's case, combined with a re-orientation away from his attachment to 'worldly' pleasures. He needed a re-orientation which provided 'an energy of a different kind, from a different source' (Murdoch 1970b: 56). Note that this kind of re-orientation is different from the kind of re-orientation most of us are familiar with. I do not only direct my own energy away from object A to object B. This re-direction may be liberating in the sense that the energy I formerly spent on attending to A now becomes available for attention to B, and I think this kind of re-direction is

involved here too. But Murdoch suggests that I also *receive* energy from B. I will say more about such reception, which will turn out to be reception not only of energy, but also of love and grace. With regard to the question of autonomy, I will explore what it would mean to go beyond the Platonic extension already developed; I want to be able to decide whether to take something on board from Augustine's model, and if so, what. For now though, I can merely reinforce the Platonic extension and stress the continuous aspect of re-direction: re-direction towards the good involves a process of 'personal history'. The difference here is that Augustine brings in God, but this difference doesn't seem fundamental for the moment. Whether we believe or not that there is an ultimate source of the good, the way we can reach the good remains the same, and – for the moment – this doesn't seem to endanger our autonomy. It is *I* who decides to direct myself to 'good' and 'God'. The mere fact that God is believed to be the source of good does not change my state of independence (outer autonomy). Therefore, this Platonic interpretation of Augustine can serve to supplement the modern ideal of autonomy in the same way as the Platonic extension: the good – divinely given – is the ultimate point of reference. Up till now, God is not decisive for the question whether I reach the good or not. This will change in the course of my argument, when I focus on other aspects of Augustine's view.

The key difference between Plato and Augustine is that the 'magnetic pole' is not 'the good' but 'God', or, at least, if the good is not directly divine, it is of divine origin. According to Augustine, God is the source of the good: 'all good proceeds from God' (II, 48). Therefore, an Augustinian extension would mean that God, not the good, is 'the one', the 'magnetic pole' we should attend to if we want to reach autonomy. The conditional status of this claim is significant. It supports my argument for compatibility between the modern ideal of autonomy and the Augustinian ideal of the person that Augustine leaves it up to *me* whether or not I turn to God. Whether or not I turn away from 'things eternal', in Augustine's words, whether or not we sin or commit evil, is a matter of free choice of my will. As I argued in Chapter 1, the wish to exercise my free will is part of the modern ideal of autonomy. Augustine's ideal of the person allows for this. I am free to turn to eternal things or not. Furthermore, just as the question whether or not my ultimate point of reference is (of) divine (origin) does not matter for my autonomy since autonomy only requires *that* there is such a point of reference; it does not matter for my autonomy whether or not my (capacity of) free will is of divine origin. The capa-

city is given to me, but I am free to do with it what I want. I can choose either good or evil. This is why, according to Augustine, God is not responsible for our evil deeds. If we choose evil, 'God the Creator is not at fault' (III, 124). If we only consider this part of Augustine's asymmetric view (that is, what he says about evil and not what he says about good), it can be integrated with the modern view that it is up to us if we choose evil, and that this freedom to choose evil is part of what it is to be an autonomous person.

Problems

The problem with this 'freedom to choose evil' view, however, is not only that in its Augustinian version it is asymmetric with regard to responsibility, as I argued in Chapter 2. The main problem, with regard to my efforts to construct an Augustinian extension of the modern ideal, is that Augustine is not able to give a good reason why we have a real choice between good and evil anyway. If I know that something is evil, and, at the same time, know 'things eternal', how can I choose evil? It is important to note that this is not a problem for Plato, since he holds that if we know the good we will also choose it. But it is a problem for Augustine, and for any efforts to construct an Augustinian extension of the modern ideal of autonomy.

Augustine attempts to deal with this problem by discussing the origin of evil, as I argued in Chapter 2. His claim that 'we do evil because we do so of our own free will' (VII.3/136) raises the question why we choose to do evil. Since Augustine does not want to consider God as the origin of evil (according to him, God is only the origin of the good), he blames the devil (VII.3/136–7) but realises that this is equally problematic since then the question is why *he* came to possess a wicked will, in other words, why *he* had any ground to choose evil given that God made him 'a good angel and nothing else' (VII.3/137).

A further problem still is Augustine's concept of grace. If grace is needed to turn myself towards the good, I am dependent on grace to reach autonomy, and therefore I no longer rule myself but God rules me. It is up to God, not me, then, whether or not I reach autonomy.

But if we want to construct an extension, why not leave out grace? Why does the ideal of the autonomous person need the concept of grace at all? Why is grace necessary for autonomy? In Augustine, I can see at least two arguments for giving grace a role in the ideal of autonomy. The first relates to inner autonomy, the second to the problem of evil.

Augustine resorted to giving a key role in the quest for autonomy to grace, I believe, because he experienced and understood how difficult it

is for persons to reach inner autonomy. 'My inner self was a house divided against itself,' (VIII.8/170) he writes in the *Confessions*, and he obviously had the wish to change his condition. But how? Consider again what it means to reach inner autonomy. It could be argued that, instead of grace, we need rather *reason* 'to do the job'. What is the role of reason in Augustine's account? I have already said that its role is similar to that which it has in Plato's account. It has an essential role in bringing about inner harmony. This is Dilman's interpretation of Augustine (and of Plato):

> As I read him, when reason is *at one* with the emotions those emotions are no longer irrational, they are shaped by the person's moral convictions [....] It is only when reason is separated from the emotions that its rule becomes despotic. The order it imposes, if it succeeds, remains external. That is not real self-mastery, for the person remains divided in himself and where he obeys 'the dictate of reason' he does so unwillingly.[25] (Dilman 1999: 77)

Harmony requires reason to be *internal*, that is, at one with the emotions, making the emotions rational and shaped by moral convictions. Only when reason is external, that is, when there is no harmony between reason and the emotions, the rule of reason is despotic. (Note that in the 'internal' case reason still rules – just not despotically.) But then there is still the problem of how to get reason to play an internal role, how to make it one with the emotions. Therefore, something 'different' is needed. This something 'different' cannot be 'moral convictions' as such, as Dilman suggests, since, as I argued before, we are able to engage in strong evaluation and question our own convictions, including moral convictions. That is why I argued for an ultimate point of reference. This then makes the rule of reason legitimate: it's 'backed up' by God (or the good), so to speak, and in that way the emotions 'accept' its rule.

But then the question arises: How do I manage to direct myself to God? In the *Confessions*, we find Augustine's testimony of how difficult this can be. Precisely because of the difficulties he experienced in his own life to reach inner harmony, I believe, Augustine concluded that we cannot direct ourselves to God without help. He had the wish to make an end to his condition of inner division, and he concluded that this was impossible without God's grace.

This, in itself, is not an argument for why the concept of grace is needed in an account of the ideal of autonomy. The fact that Augustine's

view requires the concept of grace does not imply that an account of the modern ideal of autonomy needs it. However, the need for an Augustinian extension to the modern ideal of autonomy can be made more plausible if we consider again the 'gap' identified in the modern ideal of autonomy. In the *Confessions*, Augustine writes: 'Meanwhile I was beside myself with madness… no more was required than an act of will' (VIII.8/171). What does this mean? I have already argued that rather than a single act of will there is a process of re-direction to the good involved. But it is worth considering this problem again and discussing what it means to say that by an 'act of will' I can solve the problem of infinite regress. We return here to the problem with Frankfurt's account. To put an end to the struggle between desires, a desire of a higher order is required. But where does the hierarchy of desires end? Augustine's proposal of an 'act of will' is similar to Frankfurt's 'identifying decisively with a desire' but neither solution is sufficient as it stands. To prevent the 'act of will' from becoming another participant in the inner battle of wills, rather than putting an end to this battle, we want to ensure that this act of will 'can do the job', can bring harmony in the soul. This problem leads me to consider grace. Since there is no possibility of bringing harmony from *within* – there is a battle going on – we need to appeal to help from 'outside'.

Within the Platonic framework it is argued that if we direct ourselves to 'the good', we are able to reach inner harmony, and therefore inner autonomy. But how *can* we do that, if we really are in a condition of non-autonomy, that is, in such a condition as Augustine describes as 'beside oneself with madness'? Within the Augustinian framework it is said that we need to direct ourselves to God. But this, too, seems very difficult if our will is divided. Augustine argues that the person is not able to get out of this condition by his own efforts alone; the grace of God is necessary. So in this context the difference between Plato and Augustine centres around the question of whether or not we can reach inner autonomy by our own efforts alone. This question is not only significant for an interpretation of Augustine, but also for (extending) the modern ideal of autonomy. Can we reach 'the good', in the sense of achieving inner autonomy, by our own efforts? Is it sufficient to supplement the modern ideal with a Platonic point of reference, or do we also need the concept of grace? A further question is then whether an extended ideal of the autonomous person is still an ideal of autonomy if it includes the Platonic and/or the Augustinian 'extension'. It may be that inner autonomy is reached, but does that hold for outer autonomy too? I argued that in the case of the Platonic extension there

is room for outer autonomy, since *I* direct myself to the good. If we adopt the Augustinian extension, however, it seems that outer autonomy has become impossible. If for my inner autonomy I am dependent on God and his grace, I am not independent and self-directing anymore. This problem will need addressing further.

So far I have considered an argument for including grace in the extended modern ideal of autonomy based on an appreciation of what it takes to reach inner autonomy. A further argument for the key role of grace (rather than reason alone) in the process of trying to reach autonomy, is that, according to Augustine (as I read him), our capacity for reasoning is not enough to prevent us from doing evil.

What exactly is evil? There seems to be a close connection between evil and not-autonomy. Dilman argues that 'evil is always alien to the person' (Dilman 1999: 79). This suggests that good, on the contrary, is somehow 'closer' to the person, or has at least more chance to be closer to the person. 'Closeness' means here that the person has the ability to be attracted to good and love good since there is a pre-existing link between (part of) what a person is and good, or, put in Augustine's terms, between the soul and good: (souls of) persons possess a tendency or inclination to be attracted by good. However, in Augustine's view this attraction does not prevent us from doing evil. I agree with Dilman's interpretation that Augustine argues 'that there is a part in the soul which is *drawn* to goodness' and that 'each individual soul can be awakened to the love of goodness' but still the question remains whether this 'awakening' is something we can achieve on our own or whether we need God's help in the form of grace. There is no doubt that Augustine argues for the latter, so the problem of grace and autonomy remains.

Finally, to provide a more comprehensive picture of what Augustine's ideal of the person, and in particular his concept of grace, can do and can not do in relation to the problem with the modern ideal of autonomy, we still need to consider the role of love in Augustine's ideal.

I have mentioned love already in the context of the Platonic ideal, namely as the mad love of one who beholds the beautiful and the good. This concerns the person's love for 'the one'. But Plato's ideal does not allow for love from the other direction, that is, I might love the good but the good does not love me. Even if we consider the good as related to the divine, this does not change the picture. I may be 'rapt in the divine', but the gods on Olympus *usually* don't love (particular) humans. Certainly, the ancient Greek culture allowed for much more

direct communication between humans and gods than the Christian culture did (and does). However, as the myths and phenomena such as Dionysian possession show, it is impossible to maintain that most of these communications involve love from gods towards humans. In Augustine's ideal of the person, however, love is not a one-way business. Whereas in the Hebrew texts God is shown to be demonstrating many different moods and attitudes, Christians see their God as the one who loves humans unconditionally. Augustine believes that God loves him, and that he gives him his grace.

Does this love (from God towards me), infringe on my autonomy? Consider the following argument for a negative answer to this question.

We do not generally think that human love makes autonomy impossible. Certainly, human love involves mutual dependence, but this does not exclude the autonomy of the persons engaged in this love. Perhaps it excludes absolute autonomy. Both dependency and autonomy exist in degrees. In a relationship, the degree of autonomy is lowered to allow the degree of dependence to be raised. But if a relationship of love involves total dependence, is it still a relationship of love? If I am fully prevented from being the master of my life and of the person I am, if I am completely ruled by the other person, does this person really love me? And is it worth having such a relationship at all? *Ideally*, at least, a relationship of love gives me the opportunity to be myself. If my autonomy is closely related with the search for my identity, as I argued before, then the fact that a person loves me can help me in finding and being myself, help me to achieve the inner order and harmony, help me to know what I want (my will to be), help me to reach a condition of inner autonomy.

By analogy, then, if I am engaged in a relationship of love with God, this, by itself, does not exclude but rather furthers my autonomy. God helps Augustine to be himself and to achieve inner order and harmony. The *Confessions* tell us the story of a man who, through (experiencing) God's love, manages to rule himself. Augustine's concept of sin shows that he strongly relates 'not being yourself' with what it is *not* to be autonomous: 'it seemed to be something which happened to me rather than something which I did [...]' (VII.3/136). The wish to be the author of my own actions is part of the ideal of autonomy. In Augustine's experience, only God can help us to achieve this. Augustine's search for God is at the same time a search for his own self. He wants to rule himself and be himself, and not be ruled by his first-order desires which he sees as 'not me'. Before his conversion

Augustine tells us that his inner self was 'a house divided against itself' (VIII.8/170); after his conversion he thanks God for coming to his aid and he concludes: 'Yet now I am' (XIII.1/311).

Thus, to include the Augustinian view of love in the extension of the modern ideal of the person supports this ideal by offering (as grace did) an account of *how* we can reach the ideal. Furthermore, not only is the relationship of love between a human and God compatible with autonomy. As I suggested above, inter-human relationships of love, too, are not only compatible with, they also directly enhance autonomy. I say 'directly' since they don't play a mediating role between the person and God, but are in themselves valuable to the person's autonomy. Since this does not involve dependence on God, it is an aspect that can be incorporated into the modern ideal without major problems. (Of course there is much more to be said about the relationship between human love and autonomy but I shall proceed with my main argument now.) In Plato's and Augustine's view, however, human love has rather a mediating role, enhances autonomy only indirectly. Consider the following argument. According to Plato, we can recognise the beautiful and the good in earthly things. Similarly, Augustine believes we can recognise, or find, God 'through the creatures'. Worldly things are a channel to arrive at God. Since persons are 'earthly things' and 'creatures', and – within the Platonic and Christian world-view – the most outstanding ones, to engage in love relationships with persons is an excellent vehicle for the practice of autonomy as the exercise of my capacity to achieve a condition of inner order and outer dependence through attention to the good or God.

It may be objected that although it is clear why inter-human relationships of love enhance inner autonomy, it still remains to be shown why they enhance outer autonomy. My answer is that on the one hand, certainly, I give up some of my independence; I become independent to a lesser degree. I already argued that (outer) autonomy is always a matter of degree. On the other hand, however, it makes sense to say that personal relationships of love *enhance* independence and (therefore) outer autonomy. If I see in my beloved the good, I am able to transcend my particular dependence on this particular person. Rather, I am dependent on the good as such. And this dependence does not prevent me from achieving outer autonomy but rather enhances it because of two reasons. Firstly, since this does not involve dependence on a person but on the good as such, I cannot be properly called 'dependent' on something 'outside' me since the good is in me. This argument depends on the acceptance of a controversial

metaphysics – it is even unclear whether Plato or Augustine would accept it – so I would rather leave this point aside. But secondly, the reason why being dependent on the good enhances my outer autonomy is that it allows me to evaluate myself *and* the values and norms others may try to impose on me. I have already pointed to the need for such an evaluation (Section 3.3.); the Platonic and Augustinian ideas discussed in this chapter provide a way to meet this need by offering an ultimate point of reference that allows me to evaluate societal values and norms 'given' to me and to decide whether or not I make them properly my own.

4.3. Conclusion: Overview of the argument in Part I and unresolved questions and difficulties

In Chapter 1, I set out to articulate the dominant modern ideal of the autonomous person. The wish to be an autonomous person is the wish to rule and govern oneself. I made various distinctions to clarify this ideal: the distinction between inner and outer autonomy, between capacity and condition, and between the ideal of autonomy and the ideal of doing what you want. In Chapter 2, I constructed an antithesis to this thesis. The ideals of the person and the views of freedom found in Plato and Augustine appeared to be very different from the modern ideal. However interesting they might be, it seemed that they are largely irreconcilable with the modern ideal. I argued that although there are elements in Plato and Augustine that are compatible with part of the modern ideal (in particular Plato's ideal of the person, expressed by his metaphor of the charioteer and the horses, and Augustine's ideal of the person in *On Free Choice of the Will*; both elements are compatible with the ideal of inner autonomy), most other elements found in Plato and Augustine seemed to be incompatible. Plato's view of madness, Augustine's concept of grace, Plato's reference to 'the good', and Augustine's reference to 'God', appeared to be very alien to the modern ideal of autonomy.

To reconcile ancient and modern ideals of the person, to construct a synthesis, I then first turned again to the modern ideal, the thesis. In Chapter 3, I showed a salient problem with the modern ideal. I identified the problem of infinite regress in Frankfurt's account, and considered various answers to this problem (the following elements have been considered: ideals and values (Taylor), the source of the Deep Self (Wolf), and normative standards (Feinberg)). However, these answers proved unsatisfactory, since it was always possible still to evaluate these elements, and

therefore the problem of infinite regress was not solved. I argued that the only solution was to conceptualise an ultimate point of reference that cannot itself be subject to evaluation. I concluded that the modern ideal as articulated in Chapter 1 needed to be extended with an account of the role of such an ultimate point of reference if this ideal was to be coherent. In Chapter 4, I argued that elements drawn from Plato and Augustine were able to provide such an account, and, therefore, to that degree were compatible with the modern ideal. I have shown that we can extend the modern ideal with the Platonic elements of 'the good', 'vision', 'attention', and even 'madness', or with the Augustinian elements of 'God', 'free will', 'grace', and 'love'. However, these attempts at extension have not been entirely successful. I have repeatedly shown that there are serious remaining problems. In the rest of this section I want to summarise in a different way what I have done so far and articulate the remaining problems.

One way of putting the problem with the standard model as sufficient to provide an account of autonomy shown in Chapter 3 is to say that we need some source of 'normative authority' for autonomy to be possible. Without such a source, or such a point of reference, we may *wish* to put an end to an infinite hierarchy of desires, we may *wish* to engage in strong and radical evaluation (self-evaluation and the evaluation of societal norms and values), but we won't be in a position to do this. My paradoxical sounding conclusion was that for self-rule properly understood, that is, autonomy in the 'inner' and the 'outer' sense, you need to rely on something that is not yourself. To exercise authority over yourself you need to rely on the normative authority of something that is not yourself. If this analysis is correct, then, the question is what this source of normative authority can be, this point of reference. Following Taylor, I first suggested values and norms. But this, I argued, is not sufficient, since autonomy as an ideal includes the wish to evaluate such values and norms themselves. Therefore, I concluded, the reference point or source of normative authority we want to look for is of an 'ultimate' character, something that itself cannot be the object of (normative) evaluation. Such an ultimate point of reference which can be the source of normative authority would be provided by the Platonic concept of the good and Augustine's concept of God. In this chapter I have argued for this point by considering how we might try to use aspects of Plato's and Augustine's thinking to extend the modern ideal. I say 'try', since the result remains problematic. There are at least these unresolved questions and difficulties anyone wishing to embrace such an extended ideal faces.

Firstly, modern people holding the ideal of autonomy, that is, wishing to rule themselves, are generally reluctant to commit themselves to a Platonic or Augustinian metaphysics. Even if some would be prepared to accept the existence of objective values, the idea that there is something like 'the good' or 'God' in particular is certainly not generally accepted. *If* my argument for the need to extend the modern ideal with some account of a source of normative authority is valid, it follows that it is inconsistent to hold (1) the belief in autonomy as an ideal and (2) the belief that there is no 'good' or 'God' or other source of normative authority. But is it valid? Is it really not possible to solve the problems of the modern ideal *without* resorting to such an extravagant metaphysics?

Secondly, in this chapter I have given some reasons why dependence and autonomy need not necessarily contrast. However, this remains a general problem for the (attempted) reconciliation of 'ancient' and 'modern'. If I am so much dependent on 'the good' or 'God' for my autonomy as I suggested in this chapter, how is it possible still to call this autonomy? Is the 'paradox' (as I called it above) that to exercise authority over myself I need to accept a source of normative authority that is *not* myself really no more than a paradox, or is it simply a plain contradiction? As I have shown in Chapter 1, the modern ideal includes that 'I wish my life and decisions to depend on myself, not on external forces of whatever kind' (Berlin 1997 (1958): 203). If my life and decisions ultimately depend on 'the good' or 'God', where is my autonomy?

Thirdly, I speak here of 'normative' authority. In this chapter, however, I often referred to the *moral* authority of 'the good' and 'God'. But is normative authority also moral authority? Even if my argument in relation to Frankfurt's and Taylor's model is valid, if there is an ultimate reference point needed that can provide normative authority, must this reference point also have *moral* authority? And if it has this moral authority, am I obliged to obey its authority? I will here focus on the latter problem,[26] since when dealing with Plato and Augustine, the question emerged whether or not autonomy includes the freedom to choose between good and evil. If I accept that the ultimate point of reference has *moral* authority, and I interpret this as meaning that I have no choice but to obey, I have no real choice between good and evil. If I know the good and see this as the *morally* good, if I recognise my ultimate point of reference as having moral authority, then I *have* to choose the good. This is the Platonic view. In my discussion of Augustine, however, I argued that Augustine wants to allow for the freedom to choose between good and evil. Firstly, he sees 'God' as an

ultimate reference point that has normative authority with regard to his efforts to become autonomous; it helps him to get rid of a divided will, his will becomes one. But Augustine also assigns moral authority to God. He wants God and his 'things eternal' rather than lust and the pleasures 'worldly things' can give. In Augustine's view, however, God having this moral authority does not exclude his freedom to disobey God, to turn away from him and his 'things eternal', to not accept his grace. He argues that since he has a free will (given by God), he is free to disobey. I am always still free to choose evil. But if I know God and 'things eternal', and recognise them as having moral authority, how can I choose evil? What are the grounds for choosing evil then?

This problem of the relation between autonomy and morality is not only a problem for Augustine; it is also a problem for the extended ideal I tried to develop. Should the modern ideal be extended with the Platonic or the Augustinian view of the origin of evil? Which extension is compatible with the modern ideal? Apparently, the Augustinian view with regard to the choice of evil is compatible with the modern wish to have moral freedom, the freedom to choose between 'good' and 'evil'. As I argued already in Chapter 1 (Section 1.5.6.), the modern ideal of autonomy is usually strictly distinguished from the moral question. For example, in *Harm to Self* (1986) Feinberg argues that autonomy is at best only a partial ideal, 'for since it is consistent with some important failings it is insufficient for full moral excellence' (Feinberg 1986: 45). He argues that although you may be an autonomous person, you might be cold, unloving, ruthless, or cruel – towards others and towards yourself.[27] But is such a strict division tenable? It certainly is problematic if we accept the Platonic extension. But given what I said about Augustine's view, namely that there is a problem with regard to the grounds for choosing evil, does not the modern ideal have to deal with this question too?

The three problems discussed above need further addressing. My task in Part II, then, is to find out whether there is any way to gain the advantages of the extended ideal without incurring the costs attached to Platonic and Augustinian versions. In view of the three problems discussed in this section, (some of) these costs are, firstly, a commitment to an extravagant metaphysics; secondly, the question whether and how dependence can be compatible with autonomy; and, thirdly, the question whether autonomy includes the freedom to choose evil. Can we avoid these costs? In Part II another dialectical process starts, namely between the 'thesis' of Part I as a whole, in particular the solution to the infinite regress problem presented in Chapter 4, on the one hand, and an 'antithesis' in the form of alternatives to the 'costly' (in the way I have defined costly above) solution proposed in Part I.

Part II

This page intentionally left blank

Introduction

In Part I, I attempted to reconcile the dominant modern ideal of the autonomous person with ancient ideals I articulated using Plato and Augustine. But has this attempt been successful? I have already indicated the dialectical structure of my argument in Part I. By questioning my argument of Part I, I will now generate a new dialectic, namely one between the extended ideal I constructed in Part I, on the one hand, and possible alternatives to this extended ideal, on the other hand. If we consider my argument for an extended modern ideal as a 'thesis', what would stand as an 'antithesis'? To understand the need for such an argument, let us consider again the argument of Part I.

In Part I, I approached the problem of reconciliation of ancient and modern ideals of the person from two sides. Firstly, I revealed an internal problem in the modern ideal: the infinite regress of orders of desires. To solve this problem, I suggested an 'extension' of the ideal, involving an ultimate point of reference which guides and helps the person faced with conflicting and regressing higher-order desires to achieve inner unity and harmony. Secondly, my examination of two ways of ancient self-understanding, the Platonic and the Augustinian, showed that they may be able to provide that extension in such way as to solve the regress problem. This solution, however, leaves the following three problems, as I noted at the end of Part I. Henceforth I shall refer to them as Problem One, Problem Two, and Problem Three.

Problem One

The extended ideal requires us to accept a metaphysics including the idea of the good (Plato) and/or God (Augustine). Such a commitment to non-natural entities cannot be taken for granted, and many contemporary philosophers would reject such a metaphysics. Therefore, it is

difficult to see how such an extended ideal of autonomy can fulfil the role of a contemporary dominant ideal of the person. This gives rise to the question: Is it possible to find a less metaphysically compromising extension?

In regard to terminology, hereafter I shall use the terms 'rich' and 'extravagant' when referring to the first kind of metaphysical commitments mentioned above (as with Plato and Augustine), and 'economical' or similar terms when referring to extensions of the second kind.

Problem Two

It may be objected that if I depend on something that is not me (God or good) I am no longer autonomous, since the notion of autonomy appears to involve my being independently self-determining (see for example the problem with Augustine's ideal discussed in Section 2.3.1.). A possible reply could be that if we assume a strong metaphysical relationship between 'me', on the one hand, and 'God' and/or 'good' on the other hand (for example the view that I carry a divine 'spark' or something good in me or the view assuming a (temporary) *identity* between me and God/good, as in mysticism), then a person can still be called autonomous to the extent that 'I' and that on which 'I' depend coincide. If, in determining myself, I depend on myself (alone), I am independent. On that condition, therefore, and if autonomy involves being independently self-determining, it is appropriate to call myself autonomous. However, this solution does not avoid Problem One and arguably makes it rather worse.

Problem Three

If I attend to, clearly perceive, and *know* the good or God, can I (still) act badly? Why does this present a problem? On the one hand, having the option to choose evil should I elect to do so appears to represent a key aspect of the modern ideal of autonomy, as I have argued in Chapter 1. Autonomy, on the modern view, includes having the freedom to choose, and therefore includes the freedom to choose between good and evil. It is tempting to transfer this view to the extended ideal, taking on board both the modern (Chapter 1) and the ancient (Augustine) wish to have the freedom to choose between good and evil. But such a view seems to imply that the person who has *chosen* evil is autonomous. One could try to avoid this implication by restricting autonomy to the freedom to choose between good and evil as a *capacity*. On this view, although the person is free to choose evil, if this person exercises his freedom to make that choice, then he loses his

autonomy, or does not achieve the *condition* of autonomy, since within the framework of the extended ideal to be autonomous includes being guided and directed by the good/God. Without such guidance, I become either dependent (on objects or people) or lost: I have no longer a sense of direction, no horizon. The problem with this argument, however, is that it is unclear on what basis I can choose evil – or good for that matter. Is it a groundless choice, as the existentialist idea of radical choice suggests? If not, on what basis can I choose? (I will show later (Section 8.3.) that Kant attempts to deal with this problem in his discussion of radical evil.) On the other hand, if we reject the view that I can (still) choose evil and argue (as Plato does) that when I know the good I can only act according to the good, then am I still free? And am I still autonomous? This connects again with Problem Two: If I can only act according to the good, am I not entirely dependent upon the good? And if I am entirely dependent on something outside me, am I still autonomous?

Given these three problems, I shall, in this second part of the book, look at possible alternatives to the ideal of the autonomous person with a Platonic or Augustinian extension that I have offered in Part I. I shall look at rival views that may appear to have fewer problems, either (1) by replacing or re-interpreting the ideal in a way that avoids the problems altogether, or (2) by extending the ideal in a less metaphysically extravagant way (dealing with Problem One) while at the same time securing a sense of autonomy as freedom to choose (dealing with Problems Two and Three).

The first rival view I consider is provided by Sartrean existentialism. It will be shown that the Sartrean existentialist ideal of the person avoids the three problems mentioned above, but pays a very high price for this. It follows the first strategy, namely replacing or re-interpreting the ideal of autonomy, but the result, I argue, is unsatisfactory as an ideal of *autonomy*. Furthermore, it is doubtful whether it can be considered an ideal at all. In questioning this, I shall be developing and extending the brief remarks I made about Sartre and existentialism in Part I (Sections 3.2., 3.3., and 3.5.).

Secondly, I shall look at Frankfurt's attempt to develop what can be construed as a metaphysically economical extension of the ideal. I am sympathetic to (what I interpret as) his argument against existentialism, against the ideal of having a lot of choices available (on its own), and against what I have called the 'ideal' of 'doing what you want'. However, Frankfurt's own ideal, as it emerges from his discussions of freedom, autonomy, and what a person essentially is, fails to solve the

three problems mentioned above. His concept of 'volitional necessity', his ideal of 'wholeheartedness', and his distinction between care and love, on the one hand, and morality on the other hand, are deeply problematic if they are understood as attempts to construct a viable ideal of the autonomous person. Here, too, I shall be taking further the discussion of Frankfurt's ideas which I began in Part I (Sections 1.3. and 3.1.).

How economical can we afford to be with regard to the metaphysics of autonomy? Is there a 'third way' between a 'rich' Platonic or Augustinian extension of the ideal of autonomy, on the one hand, and existentialism or preference satisfaction (doing what you want), on the other hand? A third alternative, apparently attempting precisely this, is Thomas Hill's version of a 'Kantian' ideal of autonomy. Hill traces very well various ideals of autonomy, including the existentialist ideal which is, according to Hill, less 'encumbered' than the Kantian one (Hill 1991: 30). But I shall argue that his own ideal fails to provide an adequate alternative to the heavily 'encumbered' extended ideal developed in the first part of this book.

The fourth and last alternative I consider is to extend the modern ideal of autonomy with Kant's ideal of the person found in the *Groundwork* and others of his writings. I investigate what could be understood as Kant's own attempt to extend the modern ideal in such a way as to solve its problems. His solution is worth considering for (at least) the following three reasons. Firstly, Kant was the first to put the issue of autonomy on the agenda in modern times, in particular through his introduction and use of the mutually exclusive autonomy/heteronomy dichotomy – his account has influenced the modern ideal of the autonomous person enormously. Secondly, Kant's strong view on the connection between autonomy and morality has been a target of Frankfurt's discussion. Thirdly, Kant is very aware of the problems I have identified and addresses them explicitly. His notion of autonomy, as (very roughly defined) acting according to reason, seems to avoid Problem One. Moreover, his account promises to give an answer to Problem Two (autonomy as self-rule: following the law one has given to oneself) and Problem Three (his *Wille/Willkür* distinction and his discussion of radical evil). I will discuss his achievements with regard to these, as well as identifying some remaining problems.

I turn now to consider what I identified above as the first 'rival' view to the extended ideal of the autonomous person I outlined in Part I, namely the view of autonomy provided by Sartre.

5
Sartrean Existentialism: Extreme Freedom and Groundless Choice

5.1. Introduction

The Sartrean view of autonomy is frequently referred to in the literature on autonomy, and it is also referred to as an *ideal* of autonomy (Hill 1991: 30). I have briefly commented already on Sartre's view in the first part of this book. However, within the framework of my book and at this point in particular it is appropriate to discuss it again in relation to the problems identified above because, at first sight at least, it seems to provide an attractive alternative to the extended view I have been developing. In this chapter, I will argue that the Sartrean ideal of a free chooser is not only a tempting alternative because it shares much of the stress on independence and (negative) freedom present in the (unextended) modern ideal (as expressed by Berlin, for example) but also it formulates a clear answer to the three problems indicated above. However, I will also offer some compelling reasons why we may not want to adopt Sartre's view as an ideal of autonomy.

I will first show how Sartre's arguments can be seen as a way of resolving or rather avoiding the problems I identified for the 'extended' view. I will argue that his view of freedom as the absence of constraints appeals to an important aspect of the modern ideal of autonomy. But, second, I will show that to support this claim, I have (1) left key features of the modern ideal aside and (2) detached Sartre's view from any criticism. Third, therefore, I will discuss some objections to his view. I have made objections against the existentialist view of choice in the first part of this book (Section 3.2. and 3.3.). Sartre, however, tackles some of these and anticipates the charge of caprice in his book *Existentialism and Humanism* (1946). Therefore, I will assess Sartre's arguments and show that they cannot stand for different reasons. In particular, they cannot

stand faced with the strongest objection possible against his view of autonomy, one that does not require metaphysical commitments he would be unwilling to accept.

5.2. The Sartrean view of autonomy

What is the Sartrean view of autonomy? I have already mentioned the Sartrean view of choice and freedom during my discussion of Taylor and Murdoch in Part I (Sections 3.2. and 3.3.) but I want briefly to make clear what I mean by 'the Sartrean view of autonomy' before I argue about it.

Autonomy may be defined etymologically as *self-legislation*. But what does this mean? Sartre sees the self-legislation of a person as a fact of the human condition. Since this condition is, according to Sartre, characterised by the absence of God, he argues that for man 'there is no legislator but himself; that he himself, thus abandoned, must decide for himself' (Sartre 1948 (1946): 56). What does this mean? As the title of this chapter shows, I choose to summarise Sartre's view of autonomy as consisting of the claims (a) that we have extreme freedom and (b) that our choice is groundless. I will now briefly explain the content of these claims. What is 'extreme freedom' and 'groundless choice'?

Extreme or radical freedom is, for Sartre, firstly the view that we have to choose ourselves, that we are, in some sense, the object of our own choice: 'Man is nothing else but that which he makes of himself' (Sartre 1948 (1946): 28); 'every one of us must choose himself' (Sartre 1948 (1946): 29); 'man *is* freedom' (Sartre 1948 (1946): 34). Secondly, Sartre thinks we cannot *escape* having to make choices. Sartre gives (among others) the example of a young man having 'the choice between going to England to join the Free French Forces or staying near his mother and helping her to live' (Sartre 1948 (1946): 35). It is part of the human condition, according to Sartre, that we 'cannot avoid choosing', in other words, 'what is not possible is not to choose. I can always choose, but I must know that if I do not choose, that is still a choice' (Sartre 1948 (1946): 48). Thirdly, this freedom and choice is 'extreme' in that it is also *groundless*. Sartre writes not only that 'We cannot decide *a priori* what it is that should be done' (Sartre 1948 (1946): 49); at times he also appears to endorse the view that there is no way of deciding *at all*: 'There are no means of judging' (Sartre 1948 (1946): 52).

With regard to questions of responsibility, the first two meanings discussed above entail that (1) I am fully responsible for what I make of myself and (2) I am fully responsible for all my actions since all my

actions are entirely due to my own choice. Against the objection that there may have been circumstances which prevented a person from realising himself, Sartre argues that man is only the sum of his actions. There is no such thing as 'potential genius', for example, except the one that is manifested in the actions of the person (Sartre 1948 (1946): 41). On Sartre's view, it is nonsense to claim that you're a great writer but argue that you did not write any books since you had a lack of time: if you didn't write, it's your own fault, your own responsibility. You *are* not a great writer; you are only what you make or made of yourself. We can conclude that, for Sartre, the ideal autonomous person accepts full responsibility for his actions, his life, and his person, since of all these he is the sole author. It is not clear to me, at first sight, what follows from the third meaning of extreme freedom (in terms of groundless choice) for the question of responsibility; however, an answer to this question can be constructed on the basis of my critique of Sartre's notion of commitment in relation to groundless choice later on in this chapter (Section 5.5.).

This is only a brief summary of Sartre's view of autonomy. I will discuss and critically examine his view in the next section, first, in the course of my argument why the Sartrean view of autonomy is attractive and, afterwards, in my discussion of objections.

5.3. Why we might want to adopt the Sartrean view of autonomy

The Sartrean view of autonomy may appear to be attractive for various reasons. Firstly, it appeals to an important aspect of the modern ideal of autonomy that I considered at the beginning of this book, namely the aspect encapsulated in the thoughts 'I wish my life and decisions to depend on myself' and 'I wish, above all, to be conscious of myself as a thinking, willing, active being, bearing responsibility for my choices' (Berlin 1997 (1958): 203). Sartre's ideal person is precisely that, a person conscious of himself as bearing (all the) responsibility for his choices and actions.[28] In other words, the ideal person does not have what Sartre calls 'bad faith'. Bad faith is a form of self-deception, in particular self-deception about the human condition as Sartre sees it:

> Since we have defined the situation of man as one of free choice, without excuse or without help, any man who takes refuge behind the excuse of his passions, or by inventing some deterministic doctrine, is a self-deceiver. (Sartre 1948 (1946): 50–1)

Thus, in Sartre's view this self-deception is an error, a denial of the truth about human existence. The ideal person according to Sartre can be defined then as the person who is *not* deceiving himself in this way: he does not deny that his life depends on himself and that he is responsible for his choices, but lives, acts, and makes choices in the consciousness of this realisation. The modern wish to be a freely choosing individual is radicalised in Sartre's concept of extreme freedom: we have to continuously make choices, our life *is* our choices and our actions, we are wholly responsible for them, and there is no given or pre-existing way of making these choices or resolving dilemmas we may be faced with.

This ideal is not only an ideal of freedom, but also of autonomy. If we understand autonomy as self-legislation, the Sartrean ideal of the free chooser seems to express the core of the modern ideal of autonomy, since it implies for the person that 'there is no legislator but himself' (Sartre 1948 (1946): 55–6). Furthermore, Sartre's ideal of autonomy is connected with the ideal of authenticity. Living in an authentic way, then, is to fully realise and recognise the above condition: we continuously *have* to choose, we are 'condemned to be free' (Sartre 1948 (1946): 34). It may be, as Sartre thinks, that many people do not realise this and 'deceive' themselves. However, they may (still) wish to be authentic. Does not our wish to be autonomous include the wish to be wholly and fully *myself*, realising and being fully aware who I am and what my *condition* is as a human being?

A second reason why we may want to adopt the Sartrean ideal is that it does not only seem to express the modern ideal of autonomy, it also makes the 'costly' extended ideal developed in the first part of my book redundant. The Sartrean view could be construed as an alternative to the extended ideal which avoids the three problems related to that ideal.

Firstly, Sartre need not assume the existence of 'good', or 'God'. Additionally, one of his central claims is the denial that there is such a thing as 'human nature'. This renders his metaphysics, if any, extremely economical.

In *Existentialism and Humanism* (1946), Sartre arrives at his view about human nature by drawing the consequences of the view that God does not exist. If there is not a divine creator who makes man 'like an artisan manufactures something' on the basis of a certain conception of man, an essence, then man is a being 'whose existence comes before its essence, a being which exists before it can be defined by any conception of it' (Sartre 1948 (1946): 27–8). Since God didn't conceive

us, he argues, we must conceive ourselves: 'Man is nothing else but that which he makes of himself' is therefore 'the first principle of existentialism' (Sartre 1948 (1946): 28). Now apart from this anthropological consequence of the non-existence of God, there is, according to Sartre also a *moral* consequence. Sartre finds it 'extremely embarrassing' that God does not exist, since it makes any idea of pre-existing values[29] at least problematic, if not impossible. Sartre argues with regard to the latter: 'There can be no longer any good *a priori*, since we are now upon the plane where there are only men' (Sartre 1948 (1946): 33). He concludes that therefore 'everything is indeed permitted;' man is forlorn if 'he cannot find anything to depend upon either within or outside himself' and is therefore 'condemned to be free' (Sartre 1948 (1946): 34).

We can conclude that in his view freedom and morality do *not* depend on the assumption that God, the good, or human nature exists. In other words, Sartre avoids the Problem One of the extended ideal of autonomy, namely the problem of having to assume a 'magnetic pole' or 'reference point' which can guide us in our choices.

Secondly, Sartre avoids the problem that, within the framework of the extended ideal of autonomy, I need to depend on something outside myself, such as 'the good' or 'God' as a 'magnetic pole', to make my (most difficult or crucial) choices and decisions (Problem Two). Sartre argues that, as a matter of fact, we cannot find anything to depend upon within or outside ourselves. Clearly, if this is so then the concept of guidance by 'good' or 'God' central to the extended ideal is empty and we had better replace it with the Sartrean ideal. By seeing the radical freedom of the individual as being a matter of fact in a godless world, Sartre avoids having to refer to anything the individual depends on. This could be seen as a virtue in comparison with the extended ideal, which is faced with (1) the task of settling what a 'magnetic pole' is (to the extent that this task relies on Platonism or Christianity, Problem One, the problem of metaphysical economy, surfaces again); and with (2) the burden of having to explain why autonomy and (this form of) dependence are compatible (Problem Two).

The third problem I identified with the extended view, Problem Three, concerned whether I can still choose evil. Does Sartre need to deal with this problem? His view of 'good' and 'evil' could be roughly summarised as follows. He denies that anything pre-exists as 'good' or 'evil'; rather, through my act of choosing something it is (considered as) good. But since Sartre also denies there can be *any* ground for choosing something, there is no ground for choosing evil (as there is

equally no ground for choosing good). Thus, the problem of choosing evil is avoided by saying that by choosing something it is considered as good. Sartre would not be prepared to accept a pre-existing good or ground for choosing good: it is entirely 'up to me' what I consider to be (a) good (choice). It is important to see that Sartrean existentialism may be attractive for this reason: it avoids having to give an account of the grounds for (doing) evil and confirms the person's freedom to choose (not the choice *between* good and evil but the choice to *make* something good (or evil) by (not) choosing it) as a key aspect of the modern ideal of autonomy.

5.4. What was left out

To present Sartre's view in the best possible light in the above section I have been one-sided in regard to the following. Firstly, I have disregarded most features of the modern ideal as for example expressed by Berlin. A full reading of Berlin's words as quoted in the beginning of my book reveals that indeed it is part of the wish to be autonomous that 'I wish my life and decisions to depend on myself' but that to (be able to) do so means according to Berlin (among other things) 'to be moved by reasons' since 'it is my reason that distinguishes me as a human being from the rest of the world' (Berlin 1997 (1958): 203). Furthermore, the sentence 'I wish, above all, to be conscious of myself as a thinking, willing, active being, bearing responsibility for my choices' continues 'and able to explain them by reference to my own ideas and purposes' (Berlin 1997 (1958): 203). In other words, the ideal of autonomy – as expressed by Berlin – includes, apart from the element of choice and self-direction, an emphasis on 'being moved by' and 'explaining by reference to' something (according to Berlin my reasons and my ideas and purposes). Sartre's view of freedom, as far as I understand it, excludes reference to *anything at all*. (I will back up this claim with arguments in the next section.) This brings me to my second point.

Secondly, I have silenced *all* objections to Sartre's view of freedom, perhaps giving the impression that his view of freedom without constraints (even not the world around the individual) is a tenable position, a good description of what freedom and autonomy *is* like for us (Sartre claimed to be doing phenomenology),[30] and a desirable ideal to aspire to. In the next section, I will show how problematic such a claim is by considering the most important objections to his view of freedom (understood then, as Hill does, as a view of *autonomy*).

In the former section I have left out at least the following three possible objections. Firstly, consider Sartre's argument that 'man is nothing else but that which he makes of himself' (c1) and that 'everything is permitted' (c2) follows from the non-existence of God (a).

(a) God does not exist
(b) (c1) Man is nothing else but that which he makes of himself.
(a) + ?
(c2) Everything is permitted (a),(b)

Is this a sound argument? Firstly, the premise (a) that God does not exist requires proof. To accommodate Sartre we could replace 'non-existence' with the perhaps more acceptable term 'absence', thereby avoiding having to prove that God does not exist. The premise goes then: he might exist but he is *at least* absent. But this premise too might be contested and would require proof. Secondly, with regard to his second conclusion, it is clear that the argument is invalid, that is, the conclusion (c2) that everything is permitted does not follow if the premises (a) God does not exist and (b, c1) (therefore) man is nothing else but that which he makes of himself would be true. There may be other reasons why not everything is permitted – even in the absence of God and even if we accept the ambiguous claim that man is nothing else but that which he makes of himself.

A second possible objection is related to Sartre's claim that we cannot find anything to depend upon either within or outside ourselves. This is something Sartre assumes to be *a matter of fact*. However, to say that something is a 'matter of fact' is insufficient as an argument unless evidence is given, and the argument that it follows from the non-existence of God is unsound.

Finally, note that Sartre's problematic reference to 'a matter of fact' reveals a tension between his self-claimed 'phenomenological' approach, claiming to describe what things are like for us, on the one hand, and possible counter-arguments that may claim things to be (often) quite different. Does Sartre adequately describe how we experience things (I mean: the world, ourselves, and the relationship between the two)? I will return to this objection in the next section where appropriate.

I will now further discuss objections to the Sartrean view of autonomy. It is my aim to show that in spite of the virtues considered above, his view is incapable of providing a coherent and sound account of autonomy that could function as an ideal of the person. I will have

to conclude that Sartre avoids the problems of the extended ideal of autonomy at the cost of giving up the concept of autonomy itself.

5.5. Objections

Some objections to Sartre's view of freedom I will make or refer to below may be well-known.[31] However, I doubt whether the implications of these criticisms in relation to the modern ideal of autonomy are sufficiently understood. I will discuss some objections to point out some of these implications most relevant to the project of this book. I will present the arguments in ascending order of importance, starting with objections Sartre can reasonably well deal with, proceeding to fundamental problems for his view, sketching the unsatisfactory and untenable position Sartre has got himself into.

Firstly, perhaps unsurprisingly given his view on traditional metaphysics, Sartre's view has attracted a great deal of criticism from Platonic and Christian authors. For example, I have referred to Murdoch's Plato-inspired critique of the existentialist idea of freedom in the first part of this book. Furthermore, Sartre's *Existentialism and Humanism* is explicitly aimed at addressing 'Christian' criticism. But to the extent that these criticisms are based on Platonic or Christian metaphysics, Sartre could easily dismiss them by saying that he rejects (such) metaphysics. Most objections Sartre discusses in *Existentialism and Humanism*, therefore, and despite his own claims, are not specifically 'Christian' in nature, and require a more intellectually demanding defence.

Secondly, Sartre mentions the criticism that his account of morality entails that 'everyone can do what he likes, and will be incapable, from such a point of view, of condemning either the point of view or the action of anyone else' (Sartre 1948 (1946): 24). Sartre could reply to this objection that the person who claims this is required to show why condemning or, to use a more neutral term, judging or evaluating the point of view or the action of other people is a moral requirement. Why do we need to be able to evaluate anyone at all? Sartre does not make such a reply, however, because in fact he shares the view that as part of morality we ought to be in a position to condemn (some) actions and views of others. His use of the term *engagement* as well as his actual engagement in politics strongly suggest a personal dedication and commitment to condemning others. The problem for Sartre, then, is that he needs an account of the grounds for such an evaluative activity, but within his own framework such a ground is excluded.

Thirdly, some analytic philosophers have questioned Sartre's arguments without starting off from a specific metaphysical position such as the Platonic or the Christian one. In his book *Using Sartre*, Gregory McCulloch sums up some problems with what he calls Sartre's view of 'extreme freedom'. I will focus on the problem most relevant to the question of the ideal of autonomy. McCulloch argues that to say that we are free to change even our deepest values and basic aims and projects 'underplays the fact that values and desires can have increasing degrees of entrenchment in an individual mode of being' (McCulloch 1994: 64). In other words, certain values and desires may be less important to us, whereas some aims may be part of what I am as an individual. Not all my choices are that important *to me*, that much part of *what I am*, but some surely are. McCulloch gives the example that I could abandon my plan to go out tonight, but in contrast I may not feel 'even slightly free' to give up my long-term aim of continuing as a professional philosopher (McCulloch 1994: 64). So although I chose a certain course of life, to abandon it may not be a live option for me. Indeed, 'from the fact that we can adopt a questioning attitude towards our values, nothing much follows about our capacity to see them as really changeable' (McCulloch 1994: 68). *Of course* it may be that my values change over time, but it does not follow that *at this moment* I understand myself as having a real choice or option with regard to them.

This objection, namely that something might not be a real option for me, is a problem for Sartre in that he claims to be examining our phenomenology, but his account of choice does not allow us to make sense of this particular human experience (something is not an option for me) as a *true* and *real* experience of consciousness (true and real as opposed to making an error or deceiving oneself). Whereas for Sartre all choices are equal, McCulloch's example shows that we do not (always) experience it thus. Or, to express it more strongly, we generally do not view all our choices, or our having choice, in this way. Sartre's reply to this is, famously, that we are deceiving ourselves, we are in 'Bad Faith'. But if I really do not consider a course of life as a real option for me, why should I be deceiving myself?[32] Sartre, therefore, is faced with the problem that in so far as he wants to describe the world as it is for us (being true to his phenomenological approach), he needs to abandon his view of freedom and choice. For most of us, the world is not an empty canvas on which we paint our lives; there are already certain shapes and colours that make certain moves of our brush (options and choices) appear *at least* more real or possible to us (if not excluding certain choices). Sartre needs to account for this phenomenological reality.[33]

Furthermore, if McCulloch's description is correct, we need to draw out the full implications of it for morality and autonomy. If it is indeed the case that (1) some values and aims are more important to us than others, as I argued above, and (2) that we do not want simply to take for granted this fact but direct our lives in a self-conscious and purposeful way (since we desire autonomy, autonomy is an ideal for us), then the question arises which values and aims are *good*, which values and aims *ought* we to have? That this question arises is not obvious and will receive further discussion in the next chapter, but in any case Sartre completely fails to account for this normative and moral dimension.[34] The normative question could be rephrased as a question of *choice*. Sartre could argue that we *have* to choose between values and aims. But he cannot provide a basis on which we could make such a choice (or a meta-choice if you like), since, as I will show now, for Sartre even the simplest choices are utterly *groundless*.

Fourthly, given the fact that we humans are faced with choices (at different levels, as has been pointed out above), we need to know on what basis we can make these choices. Trivial choices such as one between chocolate and vanilla ice-cream are perhaps excluded; arguably they are nothing other than a matter of taste, preference, etc. Sartre is right to point out that our lives *are* full of choices and that this can often be experienced as a burden rather than a liberation. But what choices are *not* trivial, *not* a matter of taste or 'mere preference' (whatever that's supposed to mean)? How can we differentiate between non-trivial and trivial choices? We need a criterion to do this. And if it's not a trivial matter, how am I to choose?

Sartre's answer to this question is that there is no ground for choice. Our choice is groundless. Although Sartre does not himself use the term in *Existentialism and Humanism*, I believe it appropriate to refer to his account as amounting to saying that choice is 'groundless'. I have already discussed Sartre's claim that we cannot find anything to depend upon either within or outside ourselves when making choices and that therefore we are condemned to be free (Sartre 1948 (1946): 34). This entails that in making a choice, I cannot depend on *anything at all*. It may be objected that Sartre means we only lack pre-existing, *a priori* values or principles to make a choice. But the following passage suggests that we – in Sartre's own words – do not have *any means of justification or excuse*:

> Nor, on the other hand, if God does not exist, are we provided with any values or commands that could legitimise our behaviour. Thus we

> have neither behind us, nor before us in a luminous realm of values, any means of justification or excuse. We are left alone, without excuse. This is what I mean when I say that man is condemned to be free. (Sartre 1948 (1946): 34)

Sartre could reply to my objection that I have to create my own values. But how is it possible to do this if I cannot depend on *anything* within or without myself? Is a creation *ex nihilo* possible at all? Traditionally, it is argued that only God is capable of such an action. And indeed, Sartre suggests that we have to take over God's role: 'if I have excluded God the Father, there must be somebody to invent values' (Sartre 1948 (1946): 54). But the invention of values out of the blue, by human beings, regardless of whether God exists or not, is an idea that is hard to make sense of; unless it is shown otherwise, I suggest it is unintelligible.

Sartre could reply that the invention of values simply means that 'life is nothing until it is lived; that it is yours to make sense of' (Sartre 1948 (1946): 54). This idea is perfectly compatible with the extended ideal of autonomy developed in the first part of this book, since as an autonomous being I need and want to make sense of my life, and I may need guidance to be able to do that. But the problem is that Sartre goes on to say: 'and the value of [your life] is nothing else but the sense that you choose' (Sartre 1948 (1946): 54). Then we encounter again the problem that *that* choice is equally groundless for Sartre: to choose the sense of my life I cannot depend on anything within or without myself.

The idea of groundless choice itself is perhaps not necessarily empty or meaningless, since, as I said above, we may have to make 'trivial' choices such as one between different sweets, different beers, different clothes. So we have to account for the fact that sometimes people say, when asked why they chose one rather than the other, that there was 'no particular reason', that they just felt like it. But how trivial are these choices really? Some people are very serious about their choice of beer, car, clothes, or football team. They may even claim that their choice is part of 'me', of what they are as a person. As I said before: where do we draw the line, how do we choose between alternatives that may seem equally attractive for different reasons, and how – a problem mentioned before – are we to judge the choices of others?[35] Just by referring to taste? I don't *like* that? This may be appropriate in very trivial situations, but which situations or choices are trivial?

Sartre has a problem here because he *does* assume that certain choices are *not* trivial, since he appeals to the notion of 'commitment'

and gives examples of choices that are not that trivial at all (his examples of dilemmas). If those choices *matter*, and if commitment includes being able to justify your choice and criticise the choice of others, the notion contradicts Sartre's suggestion that choices are merely expressing your preference. Sartre could give up the notion of commitment, of course; this position is available to him, but he does not want to do that. This leaves him in a very unsatisfactory position.

The more fundamental problem for Sartre's view of choice, however, is that even with regard to the most 'simple' choices Sartre is not able to provide us with a coherent account; even reference to taste is problematic. If choice is ultimately groundless, I don't even have taste to guide me. I must choose on the basis of nothing at all. But this doesn't deserve the name of 'choice'; it's caprice.

Sartre tries to answer the caprice charge by saying that 'what is not possible is not to choose' and that this is a limit to caprice since 'I bear the responsibility of the choice which, in committing myself, also commits the whole of humanity' (Sartre 1948 (1946): 48). What does Sartre mean by this? Do we, by choosing, project an image (ideal?) to the rest of humanity? Why would that make us responsible at all? And if the 'commitment' entails something more than projecting an image, perhaps something *like* choosing according to Kantian universalist ethics, in particular according to the categorical imperative, how could that ever follow from Sartre's own account? Therefore, we cannot seriously consider Sartre's reference to commitment as an objection to the charge of caprice.

5.6. Conclusion

We can conclude that Sartre's ideal of a free chooser is self-defeating since the notion of groundless choice is incoherent. Groundless choice is no choice at all. Therefore, if choice is to be part of the ideal of autonomy, the Sartrean notion of 'choice' is not a suitable candidate.

There are, however, different ideas of choice. In the next chapter (in particular Sections 6.2.2. and 6.4.1.) I look at Frankfurt's notion of 'volitional necessity', hoping to find and articulate an ideal of autonomy that includes a more adequate view of choice which nevertheless solves the problems with the extended ideal of autonomy I identified at the start in a satisfactory way.

6
Frankfurt

6.1. Introduction

In this book, I engage twice with Frankfurt's thought – the two different stages of my argument corresponding to what we can see as two different stages in Frankfurt's work. The first time was when I pointed to a principal problem in his 'early' work, that of the infinite series of higher-order desires, the 'endless regress' problem. I will show now that in Frankfurt's 'later' work (1988 and 1999) we can find a view which, construed as an ideal of autonomy, may provide an attractive alternative to the extended ideal. First I will explain Frankfurt's ideal of autonomy (constructed on the basis of his recent work) and identify its virtues; then, second, I will discuss possible objections to it.

6.2. Frankfurt's ideal

6.2.1. Frankfurt's central thesis: Love and care are essential to our autonomy

Frankfurt's central thesis in his book *Necessity, Volition, and Love* (1999) is that constraints on the will of an agent are compatible with the freedom of that agent. More boldly, he writes that 'the grip of volitional necessity may provide, in certain matters, an essential condition of freedom; indeed, it may actually be in itself liberating' (Frankfurt 1999: x). Explicitly mentioning autonomy, he writes: 'A number of my essays are devoted to exploring ways in which volitional necessities of one sort or another facilitate, or are essential to, an autonomy that they might be thought to diminish or to preclude' (Frankfurt 1999: x). In *The Importance of What We Care About* (1988) he writes:

> The notion that necessity does not inevitably undermine autonomy is familiar and widely accepted. But necessity is not only compatible with autonomy; it is in certain respects essential to it. There must be limits to our freedom if we are to have sufficient personal reality to exercise genuine autonomy at all. (Frankfurt 1988: ix)

Frankfurt does not make explicit what he means by genuine autonomy.[36] But the claim that volitional necessity is essential to autonomy is interesting enough, given the regress problem discussed in the first part of my book. If volitional necessity can serve as a term to name the limit to the problematic Frankfurtian hierarchy of desires, it could solve the problem of the modern ideal (the regress problem) without invoking the metaphysical theories of Plato and Augustine (avoiding Problem One).

Frankfurt's main claim about autonomy is that care and love constrain the person's will and, rather than impeding his autonomy, (are necessary to) make him wholehearted, augmenting the scope and vigour of his autonomy. But (a) what is the nature of this constraint (how do love and care constrain the will?); and (b) what is wholeheartedness?

6.2.2. Frankfurt's concept of volitional necessity

The answer to question (a) is provided by Frankfurt's concept of volitional necessity. Frankfurt observes that there are things and people I cannot help caring about. He argues that on the one hand, '*the will is absolutely and perfectly active*': none of my choices merely happens; they are the occurrence of *my* activity. I cannot be a passive bystander with respect to my choices. On the other hand, however, there is a sense in which the freedom of the will is subject to significant limitation. 'From the fact that there is something we cannot do passively or unfreely, it does not follow that it is an action we are always able or free to perform' and 'there may be certain choices that I cannot choose to make' (Frankfurt 1999: 80). We may be 'incapable of having that desire' (Frankfurt 1999: 80). Frankfurt cites the example of Luther saying 'Here I stand; I can do no other' to show that there are 'cases in which people do find it impossible to bring themselves to perform certain volitional acts' and this means, according to Frankfurt, that 'their wills are limited' (Frankfurt 1999: 80). This Frankfurt calls 'volitional necessity'. Frankfurt then argues that although this is necessity, it does not impair a person's freedom, 'since the necessity is grounded in the person's own nature' (Frankfurt 1999: 81).

Why does this necessity not impair the person's freedom? Frankfurt uses the example of Luther in *The Importance of What We Care About* (1988) to show that although it is often entirely up to the person what to care about, in certain instances he is susceptible to a 'somewhat obscure kind of necessity, in virtue of which his caring is not altogether under his own control' (Frankfurt 1988: 85–6). At that point he uses the term 'volitional necessity' for the first time. Commenting on Luther's declaration he writes:

> After all, he [Luther] knew well enough that he was in one sense quite able to do the very thing he said he could not do; that is, he had the capacity to do it. What he was unable to muster was not the *power* to forbear, but the *will*. I shall use the term "volitional necessity" to refer to constraint of the kind to which he declared he was subject. (Frankfurt 1988: 86)

Does this answer the question why the person is still free? I believe this question is important since the modern ideal of autonomy seems to imply the wish to be free (see Chapter 1). But what kind of freedom is this? Are there reasons that could support the claim that a person subject to (this kind of) volitional necessity is free (in some sense) and therefore autonomous?

Firstly, a person who is subject to volitional necessity and who therefore 'finds that he *must* act as he does' (Frankfurt 1988: 86) is not passive. 'People are generally quite far from considering that volitional necessity renders them helpless bystanders to their own behaviour. Indeed they may even tend to regard it as actually enhancing both their autonomy and their strength of will.' (Frankfurt 1988: 87). If this is right, in what does the 'active' role of the person consist? This brings me to the next point.

Secondly, the person does not experience the volitional necessity as something alien or as external to himself. Rather, the person actively identifies himself with certain desires (Frankfurt 1988: 87). Through identification, the 'alien' element is made 'mine'. (I will say more about this in my critique of Frankfurt's concept of wholeheartedness and identification (Section 6.4.1.).) The necessity is 'to a certain extent self-imposed' (Frankfurt 1988: 87). To this extent then, I infer, the person is autonomous, since to impose a requirement on oneself is central to the meaning of autonomy as self-rule.

Thirdly, by constraining the person 'to do what he really wants to do' (Frankfurt 1988: 88), volitional necessity helps the person to 'avoid

being guided in what he does by any forces other than those by which he most deeply wants to be guided' (Frankfurt 1988: 87). In other words, if I am volitionally constrained by my deepest cares, I am *myself*. This too is autonomy. The ideal of autonomy includes the wish to be able to do what I really want to do, to act according to what I am.

Fourthly, while recognising that 'it may seem difficult to understand how volitional necessity can possibly be at the same time both self-imposed and imposed involuntarily' (Frankfurt 1988: 88), Frankfurt resolves the difficulty by saying that 'volitional necessity may be both self-imposed in virtue of being imposed by the person's own will and, at the same time, imposed involuntarily in virtue of the fact that it is not by his own voluntary act that his will is what it is' (Frankfurt 1988: 88). In this way then, it is not only *possible* for a person 'to be constrained by a necessity which is imposed upon him only by himself' (Frankfurt 1988: 88), it is also *desirable* as an ideal of autonomy. With his concept of volitional necessity, Frankfurt manages to capture (1) the idea that autonomy is self-rule and (2) the idea that 'there must be limits to our freedom if we are to have sufficient personal reality to exercise genuine autonomy at all' (Frankfurt 1988: ix). According to Frankfurt, then, the answer to the question 'What limit does an autonomous will require?' is 'what a person cares about' (Frankfurt 1988: 110). How does Frankfurt arrive at this answer?

Related to Frankfurt's concept of volitional necessity is his distinction between ethical decisions about one's life and 'the feelings and attitudes of the person whose life is in question: what gives him satisfaction, for instance, or what he really wants' (Frankfurt 1999: 92). The latter, namely the feelings and attitudes of the person whose life is in question, is in Frankfurt's view relevant to volitional necessity: 'In order to have a basis for judging what is important to him, a person must already care about something' and this importance 'must be outside his immediate voluntary control. In other words, there must be something about which we *cannot help* caring' (Frankfurt 1999: 94). This volitional necessity, then, is taken by Frankfurt to be a 'necessary condition for making a rational choice of final ends' (Frankfurt 1999: 94). Note the way this is consistent with my argument against Sartre in the previous chapter. According to Frankfurt, the resolution of a Sartrean dilemma requires 'that he [the person] really care more about one of the alternatives confronting him than about the other; and it requires further that he understand which of those alternatives it is that he really cares about more' (Frankfurt 1988: 85). In other words, the volitional necessity of what the person cares about enables him to

make a grounded choice.[37] Furthermore, Frankfurt claims that 'an exaggerated significance is sometimes ascribed to decisions, as well as to choices and other similar "acts of will"' (Frankfurt 1988: 84), a remark which comes close to Murdoch's argument against the Sartrean conception of choice as something that goes on at particular moments rather than being part of a 'personal history'. I will say more about how Frankfurt's account supports my argument against the Sartrean ideal of autonomy later in this chapter (Section 6.3.).

By making a distinction between a person's ethical decisions and what a person really wants, Frankfurt reinforces his distinction between ethics and 'what we care about'. In *The Importance of What We Care About* Frankfurt argues that what we care about may be different from the requirements of ethics. According to Frankfurt, ethics has to do with the problem of ordering our relations with other people, and especially with the contrast between right and wrong, and the moral obligations we have towards others. What we care about may be different. We may care about personal projects, about certain individuals and groups, etc. If we suppose that a person needs to choose between alternatives, and one of them is morally preferable, there are two possibilities. Either the person does not already know which alternative is morally preferable, but does not bother to find out since 'it might be sensible for him to decline to look into the matter at all, on the grounds that under the circumstances doing so would be too costly' (Frankfurt 1988: 81). Or the person does know what he is morally obliged to do, but nonetheless chooses 'deliberately to violate this obligation [...] because there is an alternative course of action which he considers more important to him than meeting the demands of moral rectitude. It seems to me that both in this case and in the first the subordination of moral considerations to others might be justified' (Frankfurt 1988: 81). In both cases, it seems to me, the reason Frankfurt suggests why the person does not consider nor follow the ethical, is the person's identity. We are bound by

> necessities which have less to do with our adherence to the principles of morality than with integrity or consistency of a more personal kind. These necessities constrain us from betraying the things we care about most and with which, accordingly, we are most closely identified. In a sense which a strictly ethical analysis cannot make clear, what they keep us from violating are not our duties or our obligations but ourselves. (Frankfurt 1988: 91).

Since the person's deepest cares are related to his identity,[38] what a person cares about is not 'under his immediate voluntary control' (Frankfurt 1988: 85).

Note that Frankfurt's conceptualisation of volitional necessity utilises some of the terms I used in constructing the extended ideal in Part I, such as guidance, personal history, and dependency. Let me show this.

Firstly, Frankfurt views the notion of what a person cares about as coinciding in part with 'the notion of something with reference to which the person guides himself in what he does with his life and in his conduct' (Frankfurt 1988: 82).

Secondly, there is some kind of idea of 'personal history' present in his thinking about volitional necessity. As I interpret him, he argues for the idea that caring is not something that goes on only at certain moments of decision or choice, but is linked to the person as an entity with a history (constituting his identity, or, at least, a certain continuity in behaviour): 'The outlook of a person who cares about something is inherently prospective; that is, he necessarily considers himself as having a future. [...] The moments in life of a person who cares about something [...] are not merely linked by formal relations of sequentiality. The person necessarily binds them together' (Frankfurt 1988: 83). This aspect of continuity makes care different from mere desire. Frankfurt argues that desires have no inherent persistence, whereas 'the notion of guidance, and hence the notion of caring, implies a certain consistency or steadiness of behaviour; [...]. A person who cared about something for just a single moment would be indistinguishable from someone who was being moved by impulse. He would not in any proper sense be guiding or directing himself at all' (Frankfurt 1988: 84). I will return to this distinction between care and desire below when presenting the anti-Hobbesian aspect of Frankfurt's view.

Thirdly, if I care about something or somebody I make myself (more) *dependent* on this something or somebody. A person who cares about something '*identifies* himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced' (Frankfurt 1988: 83). Apparently Frankfurt does not think of this sort of guidance and dependency as being a limit to the autonomy of the person. As I said earlier in this section, Frankfurt argues that as long as the person is guided by those forces by which 'he most deeply wants to be guided' this enhances rather than impedes his autonomy.

6.2.3. The ideal of wholeheartedness

I now return to the second of the two questions I identified earlier: What is wholeheartedness? We can begin to investigate this question by considering the following claim of Frankfurt's: 'A person is volitionally robust when he is wholehearted in his higher-order attitudes and inclinations, in his preferences and decisions, and in other movements of his will' (Frankfurt 1999: 100). This does not mean that there is no inner conflict whatsoever, but rather that in case of such a conflict the person 'must be resolutely on the side of one of the forces struggling within him and not on the side of the other. […] In other words, he must know what he wants' (Frankfurt 1999: 100).

As I quoted Frankfurt earlier, his essays set out to explore 'ways in which volitional necessity of one sort or another facilitate, or are essential to, an autonomy that they might be thought to diminish or to preclude' (Frankfurt 1999: x). Is wholeheartedness, too, facilitating or essential to autonomy? Can it be part of an ideal of *autonomy*? What is the relation between volitional necessity and wholeheartedness? Frankfurt argues as follows: 'the concept of reality is fundamentally the concept of something which is independent of our wishes and by which we are therefore constrained. Thus, reality cannot be under our absolute and unmediated volitional control. […] Now this must hold as well for the reality of the will itself' (Frankfurt 1999: 100–1). Since we are not gods, 'we cannot be the authors of ourselves. […] We can be only what nature and life makes us, and that is not so readily up to us' (Frankfurt 1999: 101). Are we still autonomous then? Frankfurt argues that these observations do not really mean that we don't have a free will 'if we construe the freedom of someone's will as requiring, not that he originate or control what he wills, but that he be wholehearted in it. If there is no division within a person's will, it follows that the will he has is the will he wants. […] Although he may be unable to create in himself a will other than the one he has, his will is free at least in the sense that he himself does not oppose or impede it' (Frankfurt 1999: 101–2). Autonomy, according to Frankfurt then, is not in the first place a question of being the *origin* of your will or being in *control* of what you will, but being wholehearted in what you will. This is Frankfurt's version of what I have called the ideal of 'inner autonomy' in the first part of my book.

But what can make a person wholehearted? What can make me autonomous in this inner sense? To answer that question, we have to return to the concept of volitional necessity. We already know *that* a limit to the will is necessary for autonomy to find a grip. But *which*

limit does Frankfurt propose? I have already mentioned at the beginning of this section that love and care are, according to Frankfurt, the limit to the will and in particular the endless series of higher-order desires. Indeed, according to Frankfurt, the most fundamental limit to a person's will is 'what a person cares about, what he considers important to him' (Frankfurt 1999: 110). His caring about it consists in 'the fact that he *guides* himself by reference to it' (Frankfurt 1999: 110–11; Frankfurt's emphasis):

> About certain things that are important to him, a person may care so much, or in such a way, that he is subject to a kind of necessity. Because of this necessity, various courses of action that he would otherwise be able to pursue are effectively unavailable to him. […] These actions are not genuinely among his options. (Frankfurt 1999: 111)

What makes a person wholehearted is precisely this 'most fundamental limit to a person's will', which is, according to Frankfurt, what a person cares about. This care constrains me, and, by constraining me, unites my will, makes me wholehearted. I know what I want, since I know what I care about. In other words, I am able to direct myself (to be autonomous) through guidance by care. If this is true, the role of 'what I care about' becomes very important to my autonomy. But is it really able to fulfil this role? How strong is the constraint and 'necessity' of love and care? Is it really a necessity?

6.2.4. The necessity of love

According to Frankfurt, the necessity of love is particularly forceful and enables us to feel most truly ourselves. To stress quite how forceful volitional constraints are, Frankfurt compares them with the force of love: 'Love *captivates* us, but even while we are its captive we find that it is in some way liberating. Love is *selfless*, but it also enables us in some way to feel most truly ourselves' (Frankfurt 1999: 114). I will leave aside Frankfurt's claim that love itself can be liberating.[39] The most important argument to note in connection with the issue and character of *autonomy* is that by analogy with the role of love Frankfurt argues that volitional constraints enable us in some way to feel most truly ourselves. And in so far as 'feeling most truly yourself' is part of what we aspire to when we say that we want to be autonomous, it is an integral part of the ideal of autonomy.

Frankfurt argues that our love(s) (and in general what we care about), which enables us to feel most truly ourselves, need not be of a *moral* nature. I may be volitionally constrained by a variety of sources:

> The ideals that define the essential nature of a person need not be moral ideals, in the sense in which morality is especially a matter of how a person relates himself to the interests of others. The most decisive boundaries of a person's life may derive from imperatives of tradition, of style, of intellect, or some other mode of ambition. (Frankfurt 1999: 115)

But wherever the boundaries derive from, Frankfurt insists that what we love has the character of a 'necessity' in the sense that 'what a person loves helps to determine the choices that he makes and the actions that he is eager or unwilling to perform' (Frankfurt 1999: 129).

If this is right, how are we sure that if we feel a 'volitional necessity' we are not merely compelled (from within or without) by something that is *not* my deepest love and care, *not* something that makes me feel truly myself? What is the difference between being overwhelmed by love and being overwhelmed by a certain compulsion (for example, being overwhelmed by a drug)?

6.2.5. Being overwhelmed by love versus being overwhelmed by (other) compulsions

Frankfurt makes a distinction between, on the one hand, actions due to what I *care* about and being overwhelmed by *love* (volitional necessity), and actions due to addiction, terror, or some other overwhelming compulsion, on the other. Indeed, he says there are 'numerous emotions and impulses by which people are at times gripped so forcefully and moved so powerfully that they are unable to subdue or to resist them,' giving the examples of 'being enslaved by jealousy or by a compulsion to take drugs' (Frankfurt 1999: 136). He elucidates the difference between being overwhelmed by love and being overwhelmed due to addiction or compulsion by saying that in the latter case, we can or can not identify with or endorse these passions:

> In many circumstances we regard forces of these kinds as alien to ourselves. This is not because they are irresistible. It is because we do not identify ourselves with them and do not want them to move us. […] But irresistible forces do not invariably oppose or conflict with desires or intentions by which we would prefer to be moved. They

> may move us irresistibly precisely in ways that we are wholeheartedly pleased to endorse. There may be no discrepancy between what we must do and how we would in any event wish to behave. In that case, the irresistible force is not alien to us at all. […] Whether a person identifies himself with these passions, or whether they occur as alien forces that remain outside the boundaries of his volitional identity, depends upon what he himself wants his will to be. (Frankfurt 1999: 136–8)

This is an interesting distinction since it tells us something about the ideal of autonomy according to Frankfurt: we *wish* to be able to decide what we do with passions – we want to either see them as alien or to identify with them. According to Frankfurt, love is different:

> The fact that a person loves something does imply, however, that he cannot help caring about its interests and that their importance to him is among the considerations by which he cannot help wanting his choices and his conduct to be guided. […] It is an element of his established volitional nature, and hence of his identity as a person. (Frankfurt 1999: 137)

According to Frankfurt, being overwhelmed by love is thus different from being overwhelmed by other things, since being overwhelmed by love does not mean having been 'made to succumb in a struggle with an alien force' but rather has to do with a division of the will; in other words, rather than being overwhelmed by an external power[40] it means being overwhelmed 'by part of oneself' (Frankfurt 1999: 138). And we can be overwhelmed by part of ourselves since, according to Frankfurt, 'our essential natures as individuals are constituted […] by what we cannot help caring about' (Frankfurt 1999: 138).

6.2.6. Frankfurt's anti-Kantian argument

Frankfurt argues that an act is autonomous when performed out of love (regardless of whether the act is performed in accordance with duty): the necessity of love can provide the authority, and this authority of love must not be confused with the nature of the moral law. Frankfurt denies that it is possible to establish principles about what we ought to love: 'Love is irredeemably a matter of personal circumstance. There are no necessary truths or a priori principles by which it can be established what we are to love' (Frankfurt 1999: 130).

I read Frankfurt as meaning by 'authority' the ability to stop the endless hierarchy of desires, a sort of 'highest-order' desire, a volitional necessity. We can also understand Frankfurt's argument concerning the authority of love in the light of another way I defined the problem regarding autonomy I am trying to deal with in this book. At the end of Part I, I raised the issue of authority by interpreting my argument of Part I as an argument for the claim that autonomy requires there to be some element that has normative authority. I considered good and God as possible candidates. Frankfurt here provides something that, so it seems, can play the authoritative role without being too metaphysically costly: love. I will further discuss what Frankfurt means by 'authority' when considering objections (Section 6.4.1). For now, I continue to present Frankfurt's view and focus on his argument against Kant.

Frankfurt agrees with Kant that genuine freedom is compatible with being necessitated. However, he disagrees with Kant's construction of autonomy as submission to the requirements of duty: 'In my opinion, actions may be autonomous, whether or not they are in accordance with duty, when they are performed out of love' (Frankfurt 1999: 131). This, needless to say, goes against the heart of Kant's account of autonomy. Frankfurt argues that 'it seems natural and reasonable to presume that when a person is acting under his own control, he will guide his conduct with an eye to those things that he considers to be of the greatest importance to him' (Frankfurt 1999: 132). Frankfurt separates love and duty. Although he admits that sometimes the requirements of love and duty coincide, he writes that 'a person will not take the fact that a certain action would fulfil a duty as a reason for performing that action unless the person has a desire to do what duty demands' (Frankfurt 1999: 176). Does this mean that duty alone is not enough but needs to be coupled with desire to provide volitional necessity? Frankfurt does not explain this. Frequently, he argues that this is a question of *either/or*: volitional necessity can be provided by either duty or love, or by either *reason* or love (Frankfurt doesn't make a difference between duty and reason in this context). He does not dispute that reason can provide volitional necessity, but argues that love can also 'do the job': 'Kant insists that the requisite authority can be provided only by the necessities of reason. I believe that it can also be provided by those of active love' (Frankfurt 1999: 135)

> The lover cannot help being selflessly devoted to his beloved. In this respect, he is not free. On the contrary, he is in the very nature of

> the case captivated by his beloved and by his love. The will of the lover is rigorously constrained. Love is not a matter of choice. (Frankfurt 1999: 135)

Frankfurt is right to say that we cannot change our love by a mere act of will: 'The capacity of love cannot be entered or escaped just by choosing to do so' (Frankfurt 1999: 136). More, Frankfurt suggests we *ought* not to 'betray' what we love: 'the reason we must not betray what we love is that we must not betray ourselves' (Frankfurt 1999: 174).

6.2.7. Frankfurt's anti-Hobbesian argument

Frankfurt makes a distinction between want and desire, on the one hand, and what we care about and what we regard as important to ourselves, on the other hand. He argues that what inspires our thinking and shapes our conduct is not that we merely want something, but rather that 'we *care about it* or that *we regard it as important to ourselves*. In certain cases, moreover, it is appropriate to characterise what guides us even more narrowly by referring to a particular mode of caring – namely love' (Frankfurt 1999: 155). By making this difference, Frankfurt means to distance himself from a form of (crude) 'liberalism' (Frankfurt 1999: 156), the account of freedom provided by the view which I often summarise by the label 'doing what you want', and which, more narrowly defined and perhaps less ambiguously described than by the term 'liberalism', can be called the 'preference or desire satisfaction' view of freedom (more narrowly defined in the sense that it concerns only *getting what we want*). Both 'doing what you want' (the ideal of absolute freedom) and 'getting what you want' (the ideal of desire satisfaction) are not just taken to be an ideal of the person; often they are taken to be an ideal of society. About 'doing what you want' Frankfurt writes: 'The philosophy of liberalism is distinctively preoccupied with defining and defending the ideal of a society that maximises the freedom of its members to do what they want' (Frankfurt 1999: 156). Frankfurt objects to the 'getting what you want' ideal by arguing that the connection between *getting* what we want and actually being *happy* is very problematic. He criticises Hobbes for suggesting that 'the entire character of happiness does really lie in nothing other than the fulfilment of desire', in other words, 'what makes people happy is just doing and getting whatever they happen to want' (Frankfurt 1999: 156). Frankfurt objects to this notion of happiness since it fails to see that 'people may be misguided in what they want; he [Hobbes] takes

happiness to consist flatly in the satisfaction of whatever desires they actually have' (Frankfurt 1999: 156). So Hobbes' view is indiscriminate concerning desires. Frankfurt's answer to this problem is to make a difference between what somebody merely wants, on the one hand, and what people care about, on the other hand. I may desire chocolate ice-cream, but that doesn't necessarily mean that chocolate ice-cream 'is something that I consider to be important to me' (Frankfurt 1999: 157).

If we construct Frankfurt's account as an ideal of autonomy, then, we can include a notion of judgement, but autonomous judgement is, according to Frankfurt, not about 'ice-cream' desires but about what I consider important to me, what I really care about. It is the latter that possibly can provide the volitional necessity needed to stop the endless regress of desires and make autonomy possible.

Note that Frankfurt's distinction between 'mere' wants and what we care about parallels Taylor's distinction between two kinds of evaluation. As I argued in Part I, Taylor makes a difference between 'someone who evaluates non-qualitatively, that is, makes decisions like that of eating now or later, taking a holiday in the north or in the south' and someone who 'deploys a language of evaluative contrasts ranging over desires;' the latter Taylor calls a 'strong evaluator' (Taylor 1976: 116).

6.2.8. Conclusion: Frankfurt's ideal of the autonomous person

To conclude, Frankfurt's ideal of the autonomous person as I have constructed it is that of a person who directs himself guided by what he really cares about, what he loves. Love and care he experiences as constraints to his will, but this does not inhibit his autonomy but rather facilitates it or even makes it possible. Without love and care, he suffers from an endless regress of desires. Volitional necessity allows him to exercise self-direction properly, because it is only in virtue of his knowing what he cares about that he knows what he wants. This makes him wholehearted in the sense that his will is undivided. Although he is not himself the originator of his will, or more precisely *because* he is unable to create in himself a will other than the one he has, he knows what he wants. He is not overwhelmed by compulsion, but identifies wholeheartedly with the love that overwhelms him. This love he cannot change by a mere act of will. But it is entirely up to him whether he chooses to be true to his love or not. However, he has a good reason not to want to betray his love: to do so would be to betray himself. What he loves is what he is, it is his identity. This identity is not constituted by desires such as his desire to eat chocolate ice-cream,

but rather by what he really cares about, what he loves. And since he doesn't know all the possible consequences of his actions, he had better be careful about what he loves, he had better be careful about the person he wants to be.

6.3. The merits of Frankfurt's account

Frankfurt's ideal of autonomy has the following virtues.

1. Filling the 'gap' identified in Frankfurt's earlier account and solving the three problems identified for the extended ideal of autonomy

Frankfurt's introduction of the concept of volitional necessity can be seen as an attempt to provide an account of the constraints asked for in the first part of this book, providing a limit to the infinite regress of desires (Chapter 3, in particular Section 3.1.). It also deals with the difficult and long-standing problem of combining the ideal of autonomy as *freedom* and 'I rule me and nobody else rules I', on the one hand, with what we could call the phenomenology of constraint and *dependence*, on the other hand. This relates to Problem Two mentioned before, namely the objection raised to the extended ideal developed that if I depend on something that is not me (God or good), I am no longer autonomous. By arguing that volitional necessity does not impair the person's freedom, 'since the necessity is grounded in the person's own nature' (Frankfurt 1999: 81), Frankfurt deals with Problem Two and gets rid of the metaphysical burden of having to speak about God or good, avoiding Problem One. By discussing constraints under the umbrella of the concept 'volitional necessity', Frankfurt accounts for those human experiences which, in Frankfurt's terminology, concern a 'constraint of the will', without relying on a costly metaphysics.

The question whether Frankfurt deals adequately with Problem Three is more difficult to answer. It is not clear, at first sight, whether Frankfurt's account of love and care provides an answer to the question whether we can still choose to do evil. I can see two possible (related) answers Frankfurt's account provides. Firstly, he says that our essential nature as an individual is constituted by what we cannot help caring about. But he does not consider the question whether this 'what we cannot help caring about' is good or evil. Perhaps Frankfurt simply assumes that what we cannot help caring about is always something it is good to care about? If this is so, he avoids Problem Three, since if what I cannot help caring about is always good, then in case of a

conflict with other considerations, my deepest care takes priority. This brings me to my next point. Secondly, as I have already suggested in a note, at times Frankfurt seems to argue that there are grounds for doing evil (partly on the basis of his anti-Kantian argument). In case of a conflict between morality and what I most deeply care about, so he seems to argue, we have to choose the latter, in the sense that we are subject to the necessity of wanting to preserve our personal identity and integrity. This too seems to avoid Problem Three, but with my repeated use of 'seems' I am indicating that I shall discuss this further. For now we can conclude that Frankfurt (inexplicitly) 'deals' with Problem Three.

2. Providing good arguments against the ideal of 'doing what you want'

Frankfurt's argument that we need constraints to be autonomous is a good objection against the ideal of 'doing what you want'. I have already referred to Frankfurt's anti-Hobbesian argument above. Although in other places he does not explicitly mention 'doing what you want', he does argue against two modern ideals which constitute at least part of what I mean by 'doing what you want': (1) the ideal of having many options available (an ideal of freedom) and (2) the ideal of individuality.

Frankfurt argues that the contemporary 'ideal' of having many options available (among which I can choose) is, by itself, not desirable as an ideal of freedom:[41]

> After all, what good is it for someone to be free to make significant choices if he does not know what he wants and if he is unable to overcome his ambivalence? [...] The opportunity to act in accordance with his own inclinations is a doubtful asset for an individual whose will is so divided that he is moved both to decide for a certain alternative and to decide against it. Neither of the alternatives can satisfy him, since each entails frustration of the other. The fact that he is free to choose between them is likely only to make his anguish more poignant and more intense. Unless a person is capable of a considerable degree of volitional unity, he cannot make coherent use of freedom. Those who care about freedom must therefore be concerned about more than the availability of attractive opportunities among which people can choose as they please. They must also concern themselves with whether people can come to know what they want to do with the freedom they enjoy. (Frankfurt 1999: 102)

Frankfurt argues that, as a result of people aspiring to this ideal of freedom, we can observe 'the expansion of freedom', the 'steady and notable weakening of the ethical and social constraints on legitimate choices and courses of action' (Frankfurt 1999: 108). Thus, although the weakening of ethical and social constraints is not itself held as an ideal, it is nevertheless the observed consequence of this ideal of freedom.

'Another ideal,' Frankfurt says, is the ideal of individuality, 'construed in terms of the development of a distinctive and robust sense of personal identity' (Frankfurt 1999: 108). Frankfurt now connects this ideal of individuality with autonomy and self-determination:

> To the extent that people find this ideal compelling, they endeavour to cultivate their own personal characters and styles and to decide autonomously how to live and what to do. Insofar as men and women have attained genuine individuality, they know their own minds. Furthermore, they have formed their minds not by merely imitating others but through a more personalised and creative process in which each has discovered and determined independently what he himself is. (Frankfurt 1999: 108)

Frankfurt's critical stance, then, is that he claims that 'it is true both of freedom and of individuality that they *require* necessity', which he explains as follows:

> For if the restrictions on the choices that a person is in a position to make are relaxed too far, he may become, to a greater or lesser degree, disoriented with respect to where his interests and preferences lie. Instead of finding that the scope and vigour of his autonomy are augmented as the range of choices open to him broadens, he may become volitionally debilitated by an increasing uncertainty both concerning how to make decisions and concerning what to choose. (Frankfurt 1999: 109)

To be able to make autonomous decisions, we need constraints. Frankfurt considers the case of having no boundaries at all, the case of having every conceivable course of action available. Furthermore, the person would even be free to choose how his choices are to be made. This, Frankfurt argues, makes autonomy and self-direction impossible:

> But how is it possible for him to make that choice? What is to guide him in choosing, when the volitional characteristics by which his

> choices are to be guided are among the very things that he must choose? Under these conditions there is in him no fixed point from which a self-directed volitional process can begin. [...] Unless a person makes choices within restrictions from which he cannot escape by merely choosing to do so, the notion of self-direction, of autonomy, cannot find a grip. (Frankfurt 1999: 110)

This argument provides a valid critique of the ideal of 'doing what you want' if that ideal means unlimited choice – both in the sense of (1) having many options available and of (2) there being no limits to the choice itself, to the will if you like.

3. Providing a good argument against the Sartrean ideal of autonomy

The same argument is a key objection against the existentialist ideal of extreme freedom and absolute choice as we find this in Sartre. According to Sartre, it is always possible to choose something else; Frankfurt's argument shows that this is not only false (our choices *are* constrained) but also unintelligible. Our choices need to be constrained for us to be able to exercise our autonomy, since without constraints the notion of autonomy does not make sense. We need constraints; we need guidance, a 'fixed point from which a self-directed volitional process can begin' (Frankfurt 1999: 110).

In relation to this point about constraints, Frankfurt's account is also particularly helpful in support of McCulloch's point (Section 5.5.) that some courses of life may not even be an *option* for me. As I already quoted Frankfurt:

> About certain things that are important to him, a person may care so much, or in such a way, that he is subject to a kind of necessity. Because of this necessity, various courses of action that he would otherwise be able to pursue are effectively unavailable to him. [...] These actions are not genuinely among his options. (Frankfurt 1999: 111)

Thus, Frankfurt's account is able to deal with a phenomenon that, as I argued earlier, Sartre failed to describe adequately and give a good theoretical account of. His concept of *volitional necessity* accounts for the phenomenology that sometimes we feel we cannot do otherwise. Frankfurt interprets this as a constraint on our *will*. The problem I see with this point, however, is that both authors fail to take a further step, that is to draw out the full implications of this descriptive, phenomenological fact

for morality and autonomy. *If* it is indeed the case that some values, aims, etc. are more important to us, are the objects of our care or even love, then we can conceive of a situation in which autonomous persons do not simply want to take their cares and loves for granted, but direct their lives in a self-conscious and purposeful way. Then the question arises not only how persons *do* arrive at certain aims and cares but also which aims and cares *ought to be* more important, *are* more valuable than others. I will now argue that this question needs to be accounted for within a coherent ideal of autonomy.

6.4. Objections

Firstly, I will critically examine Frankfurt's arguments in thematic order, that is in the order I have presented them. Secondly, I will discuss objections to the claim that Frankfurt's account solves the problem of the endless hierarchy of desires (the regress problem). Thirdly, I will question whether the claim that Frankfurt's account deals adequately with the three problems the extended ideal left is justified. And fourthly, I will discuss whether Frankfurt's account of love and care can be construed as an ideal of *autonomy* at all, despite his suggestion that his essays *are* about autonomy.

6.4.1. Objections to Frankfurt's arguments (thematic)

1. Frankfurt's central thesis: Love and care are essential to our autonomy

The main critical question I would like to press is whether love and care are really essential to autonomy, are really the 'fixed point from which a self-directed volitional process can begin' (Frankfurt 1999: 110).

In the next sections I will gradually try to answer this question. The gist of my argument is that since Frankfurt does not (want to) distinguish between cares that are morally good (or right) and cares that are morally wrong, he neglects an important capacity and wish of human beings to evaluate even their deepest attachments – their own and those of others. In so far as this capacity and wish is part of the ideal of autonomy as self-conscious self-direction and in so far as we do consider our own identity – including our deepest cares – as something we can and (sometimes) wish to change, Frankfurt fails to achieve a comprehensive picture of (the ideal of) autonomy.

Note that Frankfurt's overall approach to freedom and autonomy – Frankfurt is mainly interested in the inner organisation of the will – rests on his assumption that 'volition pertains more closely than

reason to our experience of ourselves', and that it is therefore 'the more personal and the more intimate faculty' (Frankfurt 1988: viii). I believe this claim is at least controversial, and Frankfurt does not give evidence. Furthermore, Frankfurt argues that 'reason depends on will' (Frankfurt 1988: 176). However, I will disregard these remarks and rather focus on the internal problems of Frankfurt's approach and the problems related to my aim here in Part II (finding an alternative to the extended ideal that can solve its problems), rather than criticising Frankfurt's overall approach to freedom and what a person (and a person's experience) essentially is.

2. Frankfurt's concept of volitional necessity

It may be that there *is* volitional necessity, that we cannot help caring for certain things and people. But Frankfurt entirely neglects the normative question: Are all our cares morally right? Can what a person cares for really be the source of volitional necessity? And if it can be, what is the relation of this to morality? Frankfurt's account of the 'authority' of love as opposed (or next) to the 'authority' of duty is problematic. Are duty and love really that much separated, each having their 'authority', as Frankfurt suggests? Recall the modern ideal of autonomy articulated in the first chapter. Realising this ideal includes self-evaluation, including evaluation of one's desires and one's values. I do not see any reason why what a person cares for should be excluded from such an exercise. The normative question is unavoidable. Consider the following example. A committed Nazi may be 'autonomous' in the Frankfurtian sense that his will and desires are constrained by the 'volitional necessity' of his care to exterminate the Jews. However, if he is to be fully autonomous in a sense consistent with the modern ideal of the autonomous person I articulated, he is required to evaluate his care. Such an evaluation is only possible on the basis of something that is more authoritative than the care itself. This is the problem of normative authority; in Part I, I have considered 'the good' as a reference point to respond to that problem. But why should the committed Nazi choose (to be guided by) the good? Even if he knows the good, it seems that he is perfectly autonomous if he evaluates his care and chooses evil. We'll need to return to this question in relation to Problem Three, the one concerning the grounds for choosing evil. Frankfurt does not seem to deal with this question.

(For a full argument for my claim that Frankfurt needs to consider the normative question see Objection 4 below on Frankfurt's argument about the necessity of love.)

3. Frankfurt's ideal of wholeheartedness

Frankfurt briefly touches on the question concerning the source of volitional unity. He refers to Augustine's view that it is 'a gift of God' (Frankfurt 1999: 102) or that it requires 'a miracle' (Frankfurt 1999: 107), and suggests that if your will is divided, 'be sure at least to hang on to your sense of humour' (Frankfurt 1999: 107). This answer is insufficient and unsatisfactory. If wholeheartedness and volitional unity are essential to autonomy, then, if we want to make sense of our ideal of autonomy, we have to get a precise idea of how to achieve this volitional unity.

Furthermore, in relation to wholeheartedness it is difficult to see how a person being 'resolutely on the side of one of the forces struggling within him and not on the side of the other' (Frankfurt 1999: 100) can properly be said to be struggling at all. This problem could be solved by stressing the *process*, *continuous*, and *dynamic* character of wholeheartedness. The struggle may gradually disappear and wholeheartedness gradually emerge. But in this context Frankfurt does not provide for this 'process' or 'personal history' aspect of our volitional nature (he does so in his earlier work on volitional necessity though, as I mentioned above (Section 6.2.2.)).

What does Frankfurt mean by 'being resolutely on the side of one of the forces struggling within him'? Let us reconsider and re-think Frankfurt's concept of (decisive) identification and what it means to be wholehearted.

Conflicts among a person's desires are not necessarily problematic. In *The Importance of What We Care About* (1988) Frankfurt clarifies what it is for a person to 'make up his mind': 'In making up his mind a person establishes preferences concerning the resolution of conflicts among his desires or beliefs', and establishing these preferences involves 'reflexivity, including desires and volitions of a higher order' (Frankfurt 1988: 176). In other words, the person forms higher-order desires. This is Frankfurt's familiar hierarchical model I presented in Part I. However, I have also shown (in Part I) that there is a problem of infinite regress here. There may be (conflicting) higher-order desires, making wholeheartedness impossible. Frankfurt's solution to this problem is his concept of 'decisive identification' (see Part I; Frankfurt 1982 (1971): 91). We can now understand volitional necessity as enabling me decisively to identify with a desire. My deepest cares compel me to identify with one desire rather than another. Moreover, to be wholehearted, then, comes about through this decisive identification. Once I have identified with a particular desire, my will is

no longer divided. But what does this identification mean? What precisely is my attitude towards 'incongruent desires', as I shall call them, desires that I do *not* (want to) identify with? What do I do with them? What do I have to do with them if I want to single out the one I identify with? Frankfurt admits that the notion of 'identification' he used in his earlier work (quoted above) is 'terribly obscure' (Frankfurt 1988: 167). Consider therefore the following argument.

Firstly, Frankfurt's concept of decisive identification suggests that when I manage to identify decisively with one desire, the other desires are still there, they still exist. This is far from obvious. If what I really care for is really constraining my will, constituting a real volitional necessity, then why would I want *anything else*? Why would there still be any other desire at all? If this desire is *me*, there is no reason why I, aspiring to be an autonomous person, want to have any other desire that is not me. Therefore, *extermination* of 'incongruent' desires, desires not congruent with the desire I identify with, seems to be the optimal solution.

Secondly, however, there are good reasons why this solution might be optimal but not final or not possible. Even if we want to exterminate incongruent desires, our experience tells us that there is a fair chance that they will resist or return. Furthermore, Frankfurt's ideal of the wholehearted self does not need to assume complete absence of other desires or even of conflicting desires. As Velleman notes in his essay 'Identification and Identity' (2002): 'A person can be wholehearted in Frankfurt's sense while retaining desires that conflict, so long as he has decisively identified with one of the desires and dissociated himself from the other' (Velleman 2002: 100). Frankfurt argues in his later work (when he *does* say more about his concept of identification) that there are two different sorts of conflicts between desires. One sort concerns a competition for priority or position in the hierarchy of desires. The other sort concerns the question whether the desire should be given a place in the hierarchy at all. If the first sort of conflict is resolved, the competing desires are integrated; they are given a specific position. Resolving a conflict of the second kind, however, involves 'a radical *separation* of the competing desires, one of which is not merely assigned a relatively less favoured position but extruded entirely as an outlaw', since 'a person's autonomy may be threatened even by his own desires' (Frankfurt 1988: 170–1). I see at least two problems with this account.

Firstly, Frankfurt does not tell us how we are to decide when to extrude a desire or when to give it an order in the hierarchy. I guess the

criterion is whether it is indeed a desire that threatens my autonomy, that is, whether I can identify with the desire or not. But to explain this requires more work. Secondly, if the process of identification involves the radical separation of a competing desire, it is highly questionable whether this really makes me wholehearted and is attractive as an ideal. Extruding a desire as an outlaw does not eliminate the conflict; rather it is transformed into a conflict between the 'outlaw', on the one hand, and the person who has identified with the rival desire, on the other hand. The unwelcome desire is expelled from the self. Velleman claims that 'this prescription of self-health is undeniably attractive. The question is whether it attracts us by articulating what would in fact be ideal for us, given how we are constituted. I suspect that it attracts us for other reasons' (Velleman 2002: 101). Velleman argues that in fact this 'ideal' is what Freud would call 'repression'. And since repression causes, or can cause, neurosis, it is highly questionable whether this is an ideal at all. To dissociate oneself from an emotion may make the person ill. It may be objected that this analysis draws heavily on Freud, and that Freud's theories are highly controversial. However, as Velleman points out, 'beneath the theoretical apparatus of Freud's account lies a piece of folk wisdom about dealing with mixed emotions. When we are angry with someone we love, the first step toward dealing with our anger is to let it mingle with, and be modified by, our other emotions toward the same person. Isolating our hostility from our other feelings is a way of not dealing with it, of allowing it to remain undigested, a lasting source of inner strife and outer impulsiveness' (Velleman 2002: 103). Surely this 'mingling' might be very difficult and even terrifying. However, to let a desire into our emotional life may be the 'only chance of domesticating it' (Velleman 2002: 104). Therefore, Velleman concludes, Frankfurt's ideal tells us less about the constitution of the self than about our own wish to be wholehearted.

However, if we accept Velleman's critique, or at least the insight that repressing a desire may potentially cause illness rather than health, we do not need to drop the concept of identification altogether. We could still say that it is part of exercising autonomy that the person identifies with one desire more strongly than others. However, we could argue that ideally the other desires should not be repressed, but rather 'domesticated', to use Velleman's term. What is the meaning of domestication in this context? To domesticate something is quite different from expelling it. The foreign, alien element is still recognised as alien *first*, but then, instead of expelling it, it is let into the *domus*, the

house. It becomes familiar with us, and we become familiar with it. At the end, it is recognised as 'ours'. (Compare this with the domestication of an animal by human beings.) This process can also take place within the person. The house is here the self. After identifying the desire, we let the desire enter our self and make ourselves familiar with it. Finally, it will be 'ours', not in the sense that we see it as 'the best in us', but we still recognise it as part of us. From the side of the desire, there is a process of domestication which makes the desire increasingly less 'wild'. There is control, but this control is not like repression, but much more like 'harmonisation'.

Note that this concept of domestication is not only useful to criticise Frankfurt's account as an alternative to the extended ideal of the person developed in Part I, it is also helpful to clarify that extended ideal. Consider again the Platonic concept of order and harmony within the soul (and Dilman's interpretation of that concept). The Platonic ideal of inner harmony which has been made into a distinctive and defining feature of inner autonomy could be constructed as the result of a process of the domestication of the black horse by the charioteer reason on the basis of his vision of the good. First, the black horse is wild. It wants to go in all kinds of directions, and usually not the same direction as the other horse, the white horse. This is the condition of anti-harmony and anti-autonomy, the not-ideal of the innerly divided self all the authors discussed and (most of us) want to avoid. Now to domesticate the black horse is quite different from repressing it. Repressing it would mean that I try to neglect the caprices of the black horse. I try to not pay attention to it. I'm afraid of it. But at the same time I hate it. And I hate my hate. I hate the black horse and I would like to cut the reins and get away from it. But this is not possible. (And even if I would or could do that, it would come back.) I will remain innerly divided. The only way to 'deal' with the desire I do not want is to accept it as part of my self. If I accept the black horse as part of myself, I can start to try to domesticate it. I, the charioteer, will use my whip. The whip of reason, finally, will domesticate the black horse, and let it run in harmony with the white horse. My self is not divided any more, but is ordered and in harmony.

The black horse stands for an 'unwanted desire'. At first, this may seem paradoxical: if I have a desire, I desire something, I want something. In which sense is it 'unwanted' then? It is a merit of Frankfurt's account that we can understand this as a conflict between a first-order desire (I want X) and a second-order desire (I do not want to have this first-order desire to X). Now to solve this conflict and restore harmony

within the self two solutions are considered in the metaphorical account above. Frankfurt's solution is not to identify with the unwanted first-order desire. This we can understand as repression. I hate my desire, I want to get rid of it. But this boils down to the formation of a higher-order desire not to want the first-order desire. The conflict remains, and is rather intensified. Therefore, the alternative solution proposed here is to identify the unwanted desire in the sense of accepting it as part of what you want and what you are. This then allows domestication, understood here as the use of reason to harmonise the conflicting desires. I accept that the first-order desire is part of what I want, but I have a good reason not to endorse it. If I were to rely on will alone, the problem would remain unsolved. Making the point that only a non-volitional element such as reason can help out here to bring about harmony, then, shows Frankfurt's failure to rely on the person's volitional structure alone to solve the regress problem.

For the extended ideal of the autonomous person, this conclusion means that a more coherent account is given of what it means to have the capacity and be in the condition of autonomy. We see that self-rule includes self-control in the sense of the domestication of unwanted desires, desires I do not like to identify with. Now it has been part of my argument in Part I that to be able to exercise this capacity and to reach this condition of inner autonomy, I need the basis of a vision of the good or a vision of universal reason. This point also makes sense in connection with the concept of domestication. My vision of the good and/or of what is required from me by universal reason will guide me, and help my 'charioteer' in the domestication of the 'black' elements in my self. If I don't know where I'm going, how can I expect to have control over my horses? How can I bring order and harmony in my house (my self) without an idea of what is good and reasonable?

4. Frankfurt's argument about the necessity of love

The problem I have with Frankfurt's concept of the 'necessity' of love is the same as the one I have with his concept of the 'necessity' of care. Frankfurt admits that 'since people are often mistaken about what is moving them in their choices and in their action, they may be also mistaken concerning what they love' (Frankfurt 1999: 130). But it is perfectly possible to imagine a case which involves a person *not* mistaken about what he loves but who *nevertheless* loves something evil. The committed Nazi may know that he loves killing Jews, that this is his deepest care which constrains his will, but this doesn't make his love *good*. Frankfurt would accept that if the *consequences* of his love are

evil, he shouldn't *do* what he *does*, but Frankfurt provides in his account no way of judging the Nazi's *love* itself. He simply takes it for granted as a 'volitional necessity'. Frankfurt considers only the 'is' and not the 'ought'.

This is-ought problem can be put again in terms of authority. Frankfurt rightly observes that (our deepest) love(s) and care(s) *do* play an important role in our decisions, our choices, and our life. But the question I want to press in this chapter is whether they (always) *ought* to play this role, in particular whether they have any *normative*[42] authority, as opposed to merely motivating authority.

But why need Frankfurt consider the 'ought'? My main argument against Frankfurt here is not that we need to evaluate our own cares because we 'ought' to. This would amount to saying that we need to evaluate our own cares since we need to evaluate our own cares and love: this is begging the question. The reason why Frankfurt is wrong to see love and care as *given* is that there is good evidence that persons do not always see their love and care nor those of others as given. My argument is this.

Firstly, if persons have the *ability* (though they may not always use it) to question their deepest love and care and that of others, we cannot assume that love and care are given and established. Secondly, there is good evidence that persons have this ability: persons do sometimes engage in deep self-evaluation (see Taylor's concept of deep evaluation) and the evaluation of the deepest love and care of others (with regard to my example of the committed Nazi given above this means that whether or not he evaluates his care, and if he does so, whatever the outcome of that evaluation may be, *we* evaluate what he cares for). Thirdly, it follows that our deepest love and care is not to be considered as simply given and established. Fourthly, if our deepest love and care is not given and established, it is not able to fulfil the role Frankfurt gives it, namely, the role of constituting a volitional *necessity*. Fifthly, if (our deepest) love and care does not constitute volitional necessity, it also fails to stop the endless regress of desires. Sixthly, if Frankfurt's account of the volitional necessity of love is not a good candidate to solve the infinite regress problem and if it does not make sense of our ability to evaluate our own deepest love and care, it cannot be the best possible ideal of autonomy, since we would not want the ideal autonomous person to suffer from an infinite regress of desires and deny him the possibility of deep self-evaluation (weak conclusion). Finally, Frankfurt's ideal of autonomy is not an ideal of autonomy (strong conclusion).

Why suddenly this strong conclusion and what reasons support it? Consider the weak conclusion. At first sight it appears relatively unproblematic, given the argument that precedes it. But the problem lies in the added claim at the end. Are not the absence of infinite regress and the possibility of deep self-evaluation essential to autonomy itself? If so, then Frankfurt's ideal is not only *inferior* to the best ideal of autonomy but simply not an ideal of *autonomy*. An argument for this strong conclusion, therefore, has to prove that (1) the presence of infinite regress and (2) the absence of the possibility of deep self-evaluation destroy the possibility of autonomy. I regard (1) as proven in Part I of my book. How am I supposed to be self-directing and self-governing if there is no limit to my (higher-order) desires? Claim (2), however, needs more work. Why is deep self-evaluation essential to autonomy? I believe the answer to this question lies in the relationship between autonomy and identity, which I have discussed already to some extent above, in particular in the context of my former objection (aided by Velleman). Let us consider Frankfurt's argument again in the light of the issue of identity and autonomy.

According to Frankfurt, the fact that a person loves something is 'an element of his established volitional nature, and hence of his identity as a person' (Frankfurt 1999: 137). But how well 'established' are this volitional nature and this identity? Frankfurt is right to say that 'our essential natures as individuals are constituted, accordingly, by what we cannot help caring about' but he fails to account for another 'fact', namely that we as persons are able to question this 'what we cannot help caring about'. It may be that I cannot change my identity easily, but, to the extent that what I care about constitutes my identity, there isn't something like a fixed identity that provides the last and ultimate constraint to my will. Why? As I argued above, 'what we cannot help caring about' is not always able to really stop the regression of desires. I may be unsatisfied with 'what I cannot help caring about'. Therefore, in so far as 'what I cannot help caring about' constitutes my identity as a person, I can question my identity. And is not this questioning of my identity the ultimate expression of autonomy? If Berlin is right about the meaning of autonomy, namely that it has to do with being 'self-directed' and being 'conscious of myself as a thinking, willing, active being' (Berlin 1997 (1958): 203), then to question my (deepest) self is self-direction in the highest degree, and to question my identity shows me as being extremely conscious of myself.

Since this questioning relates to my identity, the person I am, questioning my deepest cares can be a disturbing thing. Therefore I may

need guidance of some sort. Frankfurt claims that love 'guides him [the person] in supervising the design and the ordering of his own purposes and priorities' (Frankfurt 1999: 165). But who or what is 'supervising' or 'guiding' his love?

5. Frankfurt's distinction between being overwhelmed by love and being overwhelmed by (other) compulsions

Frankfurt makes a distinction between being overwhelmed by love and being overwhelmed by other things: the latter means having been 'made to succumb in a struggle with an alien force', whereas the former means being overwhelmed 'by part of oneself' (Frankfurt 1999: 138). Again I can object: How fixed is this 'oneself', how fixed is our volitional identity? And if we are overwhelmed by part of ourselves, are we really overwhelmed at all?

Furthermore, if being constrained by love is different from addiction because I identify with it, what is this identification but the taking of an attitude to my love? In other words, we *are* able to question our love. This supports my argument above (objection 4) against Frankfurt's view that our love is *given* and our volitional nature *established*.

If this is true, then Frankfurt's view of the relationship between autonomy and love lacks coherence. In *The Importance of What We Care About* (1988) Frankfurt suggested that when we love something, our relationship towards what we love 'tends towards *selflessness*' (Frankfurt 1988: 89) and describes the person's experience as follows:

> His attention is not merely concentrated upon the object; it is somehow fixed or seized by the object. The object captivates him. He is guided by its characteristics rather than primarily by his own. Quite commonly, he feels that he is overcome – that his own direction of his thoughts and volitions has been superseded. How are we to understand the paradox that a person may be enhanced and liberated through being seized, made captive, and overcome? Why is it that we find ourselves to be most fully realised, and consider that we are at our best, when – through reason and through love – we have lost or escaped from ourselves? (Frankfurt 1988: 89)

Am I really selfless when I love? I might be less self-centred; the object of my love captivates me and holds my attention. But this does not mean that I lose my self. On the contrary, to be able to be 'captivated' there has to be a self.

Furthermore, if we are able to question our love, Frankfurt's description of the experience of love is not a description of a volitional necessity that enhances autonomy. Autonomy includes the (exercise of the) capacity to evaluate one's love. The exercise of (strong) evaluation is the criterion that distinguishes between autonomous and non-autonomous experiences of 'being taken over'. Whether or not the experience described by Frankfurt is *liberating*, if I am captivated, seized, and overcome by love without exercising my capacity to evaluate this love, I am not *autonomous*.

Strong evaluation of love includes *moral* evaluation. Consider love of Dionysus, for example. If a daemon or god takes possession of me, and this means that I (temporarily) lose my capacity to judge my actions (for example killing someone), then regardless of whether this liberates me or not, I am – in this state of possession – not in a condition of autonomy. If, on the other hand, and to give an example of *non-moral* evaluation, I feel that my love of a particular person or God reveals to me who I really am, enhances the exercise of strong evaluation, helps me with the evaluation and construction of my identity, and furthers inner harmony and order within my volitional structure (to recall my concept of 'inner autonomy'), then this enhances my autonomy.

It may be difficult to decide whether my love is good or evil, or whether or not my love enhances my autonomy. It makes sense, therefore, to ask whether within Frankfurt's framework there is any guidance available to the person. I will return to this later in this chapter (Section 6.4.2.) when I say something about whether Frankfurt deals with Problem Three of the extended ideal.

6. Frankfurt's anti-Kantian argument

Frankfurt is right to suggest that there are many problems related to the concept of duty and duty ethics in general. But it does not follow from this that love can be given an 'authority' without much further qualification. What does Frankfurt mean by 'authority' (Frankfurt 1999: 135, 137, 139)? (See also Section 6.2.6. and my criticism of volitional necessity earlier in the current section.) What does it mean to speak of 'the authority love has for us' (Frankfurt 1999: 139)? Is it *moral* authority? I argued before that not all love is morally good, so why then should it be given moral authority? And if it is 'motivational' authority (Frankfurt 1999: 137) what does that mean? That my will is constrained by it? Frankfurt seems to mean that love has the 'authority' to end the potentially endless hierarchy of desires. In that sense, then, Frankfurt can say that autonomy requires such an authority

(Frankfurt 1999: 135) since for autonomy to be possible the endless regress of desires has to be stopped. But does love have such an 'authority'? I have argued above that it has not even the *possibility* to do so since we are able to question our love and even change it. Constraint is not the same as authority. I may be constrained by what I love, but I may deny this constraint the 'authority' it 'wants' to exercise over my will and try to liberate myself from it (to take up Murdoch's thought again: I may re-direct myself, acquire other objects of attention, etc.). For example, perhaps I 'cannot help' loving someone, this love constitutes *here and now* a volitional 'necessity' and tries to exercise 'authority' over my will, but I may decide to struggle against this and try to think of something/someone else. In other words, there is no real volitional 'necessity' nor motivational 'authority' here. In relation to Frankfurt's argument against Kant, his claim that the authority can be provided by reason *and* love (Frankfurt 1999: 135) has to be rejected. We need not deny the importance of love in our lives and to the identity of our person, and the *quasi*-necessity it may constitute to our will at certain times, to insist upon the possibility of exercising our freedom to question or resist our love – up to the point of questioning our own identity, as I argued before. Indeed, it may be that to deny the 'authority' of my love, I have to change myself and accept the consequences. But it is entirely up to me whether I do this or not. Is this not part of what it is to be autonomous? Autonomy does not only include self-direction in the sense that you steer your car one way or other, choosing one path in life rather than another. It may also involve changing *that which is being directed*, in other words, yourself. (And since this is not a light matter – 'Who am I to be?' is a very difficult question – I suggest it had better be *guided* self-direction, bearing in mind the argument in Part I of my book.)

Although Frankfurt holds on to the 'authority' of love in relation to volitional necessity, apparently he allows for some notion of what I choose to call *resistance* to the authority of love. Frankfurt argues that just as the theoretical necessity of duty cannot ensure that rational agents will always be virtuous, volitional necessity 'cannot ensure that lovers will be true' (Frankfurt 1999: 141). This suggests that Frankfurt sees it as ideal that all lovers are always true. But there may be moral situations which make this ideal unworthy of admiration. Luckily lovers are not always true: if somebody truly loves Nazism I'm very happy when this person is not true to himself. Furthermore, there may not only be situations and consequences that make this ideal

questionable; my love itself may be evil. Therefore what we could call *full obedience* or *submission* to the 'authority' of love cannot be an ideal of the person. Furthermore, to be always obedient to authority can never be an ideal of *autonomy*. If Frankfurt wants to save (what I call) his concept of resistance he needs to allow for *judgement*: I have to decide when to obey the authority of love and when not; in other words, I have to decide whether or not my love is *good*, and whether or not it is *good* to yield to its yoke.

Frankfurt assumes a sharp distinction between love and reason. What does Frankfurt mean by love? According to Frankfurt 'the will of the lover is rigorously constrained. Love is not a matter of choice' (Frankfurt 1999: 135). But is love *never* or to *no degree whatsoever* a matter of choice? Would it not be more helpful to make a distinction between on the one hand what Kant calls 'pathological love', which, although I cannot control it *directly*, I can take at least a stance towards, I can say that I like it or not, I can 'endorse' it or not? And, on the other hand, love for the good, which is something that, once I 'have' it, I am bound by (I cannot freely choose to turn away from the good once I recognise it is good), and perhaps gladly bound by (unless the true vision of the good is something difficult to live with…)? Frankfurt is wrong in calling the commands of love 'categorical' if he means that 'pathological love' is categorical. However, he is right to say that we cannot change our love by a mere act of will: 'The capacity of love cannot be entered or escaped just by choosing to do so' (Frankfurt 1999: 136). But here Frankfurt stops. Murdoch would say that we are able to shift the focus of our attention and acquire new objects of attention. Of course, it may be that to be able to do that we need guidance by 'good' or 'God'. However, Frankfurt is (and so too am I, at least in my effort to look for alternatives to the extended ideal) unable to take this course, since this would involve again a costly metaphysics. (But perhaps not all forms of guidance bear such a cost? In the next chapter, we will look at Hill's effort to develop a Kantian ideal of autonomy which is more or less metaphysically economical.)

Is Frankfurt's sharp distinction between care and love, on the one hand, and morality, on the other hand, tenable? Frankfurt makes a distinction between caring for something and 'judging it to be valuable' (Frankfurt 1999: 158). The latter pertains to whether a person 'is *committed* to his desire for it' (Frankfurt 1999: 161). But again, what about a committed serial killer? If Frankfurt's account of love and care is to say something about autonomy, he is misdescribing autonomy. Autonomy as an ability includes the ability of judging our cares. His

description of volitional necessity gives a one-sided picture of human experience. Frankfurt gives an accurate account of the experience a person has when he feels he 'cannot do otherwise' since he loves and cares for something/somebody, but fails to account for the fact that we are able to (and sometimes do) evaluate that love and those cares. We are able to judge them. I can ask the question: Is what I care for (really) valuable? Furthermore, if the result of our evaluation is that we want to change what is most important to us, we are able to do that (even if it may take a lifetime). We can not only evaluate but also *change* our commitment. We can, ultimately, change our identity. The modern ideal of autonomy includes that wish, namely the wish to at least have the possibility to engage in that sort of activity (deep self-evaluation). We want to be able to decide who we are. Perhaps we want too much; it is conceivable that we over-estimate our possibilities to carry out deep self-evaluation in a *radical* way. We may be wrong to assume that we can *fully* decide who we are. But to deny any (modest) wish on the part of the person to engage (to a certain extent) in deep self-evaluation, to ask whether the 'is' (what I care for) corresponds with the 'ought' (what I ought to care for) is to deny the person the exercise of his autonomy as a capacity. Therefore Frankfurt's sharp distinction between love and judgement, corresponding to that between identity and morality, does not provide an adequate account of the modern autonomous person and cannot be part of a modern ideal of autonomy.

Frankfurt's anti-Kantian argument includes the argument that 'a person will not take the fact that a certain action would fulfil a duty as a reason for performing that action unless the person has a desire to do what duty demands' (Frankfurt 1999: 176). But Kant's point is precisely that we ought to and can act from duty, even if this means that we do our duty *ungern*, without a desire to do it. Frankfurt's view, if construed as a normative framework, leaves room for any desire to get in the way between me and my duty, as long as this desire is connected with 'what I really am' and 'what I really care for'. But, as argued above, what I really care for may be not *good*. The point is, again, that I can always ask the question whether my deepest cares are those I want to have. Therefore they cannot provide volitional necessity. And since autonomy requires volitional necessity, and love and care are not the right candidates to constitute volitional necessity, love and care are *not* to guide the process of self-direction I engage in as an autonomous person. His divorce of autonomy from morality is, first, problematic if it is to provide a *full*, *complete* (descriptive) picture

of autonomy and if it is to function as the *best possible*, *richest* (normative) ideal of autonomy (for the argument see also my weak claim above, objection 4); and, second, it is highly problematic if it is to function as an ideal of *autonomy* (for the argument see also my strong claim above, objection 4).

Perhaps the reason why Frankfurt makes such a strict distinction between morality, on the one hand, and love and care, on the other hand, is his inadequate view of the nature of morality. However, I choose not to elaborate this remark about morality and rather focus – as I did above – on his misinterpretation of *autonomy*.

7. Frankfurt's anti-Hobbesian argument

Frankfurt is right in his objections to Hobbes. But just as Hobbes is indiscriminate concerning desires, Frankfurt is indiscriminate about what people care about. He is right to say that a person 'undertakes to guide his conduct in accordance with what he really cares about' (Frankfurt 1999: 114) but does not consider the normative question whether those cares may be good or right. As said before, Frankfurt takes our cares as *given*. This neglects an important aspect of persons, namely the capacity and desire to subject our cares – even, sometimes, our deepest cares – to our own critical evaluation, to using our practical reason to judge whether our deepest attachments are the ones we want to have. Frankfurt defines a person in volitional terms, so if we put this criticism in his own terms, we could say that he (a) gives a very good account of what it is for us to will and to will the will we want etc.; (b) rightly points out that there is a constraint to this will; (c) perhaps even correctly says that this constraint is provided by our love and care; but then (d) fails to see that this love and care is unable to fulfil the role of a *final*, *ultimate* constraint, that love and care can itself be subject to evaluation, such that they are not just a 'given' resulting from contextual, social, personal historical *causes* but something we (sometimes) may *wish* to question. (Maybe Frankfurt does not realise this; or, if he does, does not want to consider it, since he stresses the volitional and plays down the role of reason.)

Frankfurt is right that we may be guided by moral ideals derived from tradition, etc. but, as this book shows, it is at least *possible* to question these ideals as a person and as an exercise in philosophy – regardless of whether it can be done successfully or whether it can have any influence on changing these ideals. Moreover, it is part of being autonomous to see traditions as not producing imperatives or 'necessities' leaving us no choice. I can judge and reflect upon them, I

can make them part of my person. An ideal is not really my ideal if I have not made it my own; to be autonomous involves engaging in 'strong evaluation' (see Section 1.4., where I discuss Taylor). I need to consider if what I care for is something it is right to care for, that what I attach importance to merits this attachment. And, to do so, it is necessary that I be guided by something or someone.

It may be objected: Why is it necessary that I judge myself and others? Why should I engage in strong evaluation? But part of what it is to be autonomous *is* to be committed to making strong evaluations, as I argued above, and therefore *if* I care to pursue an ideal of *complete* autonomy, 'perfect' autonomy so to speak, then I also have to aspire to ask whether my deepest cares are the ones I ought to have. As noted already, this might be a difficult question to ask. But nothing in this book has suggested that autonomy is an easy or straightforward ideal to strive for: on the contrary. But, as I *also* suggested in the first part, there might be ways I can avoid getting disorientated by seeking appropriate guidance. However, the problems related to my filling in the content of this guidance, by pointing to the guides 'good' and 'God', have generated serious problems.

Note that self-examination in the strong normative sense dealt with here is a private and personal matter in the sense that only I and I alone can engage in this activity of strong evaluation of my deepest cares and (therefore) my identity. But it may be that what can guide me in such an evaluation is not a private matter at all.

Note also the following qualification. To present a complete picture of Frankfurt it is fair to say that at one point he makes room for judging our love and care, namely evaluating them in terms of the consequences of our actions. Frankfurt argues that we need to 'be careful to whom and to what we give our love' (Frankfurt 1999: 173). According to Frankfurt, the reason why we need to be careful about what we love or care for is not that there are moral principles, or that there is morality as such, but rather because of the possible consequences of our actions. And seeing the consequences of your actions is, as I interpret Frankfurt, a problem of knowledge and power. If we were omnipotent agents, he argues, we would have nothing to fear, nothing could happen to us. But 'in view of the harms to which loving exposes us' (Frankfurt 1999: 173), we need to be discriminate in what we love. In fact, by saying that we need to be discriminate in what we love, Frankfurt undermines his own argument that love constitutes a volitional 'necessity' and 'authority' that is of a *given* and *established* nature. If we take this particular argument seriously, we

have to conclude that, according to Frankfurt, we *are* to judge our love (on the basis of the consequences). Therefore we can, on this account, take a stance on our love, evaluate it, and change our will and actions accordingly. This appears to me right, except that there may be other reasons why we need to be discriminating in what we love than the consequences of our actions and in particular the fact that we can foresee them. Frankfurt seems to hold a very narrow and superficial view of morality.

6.4.2. Consequences

What follows is a summary of the objections made and their consequences for (1) the problem of the endless hierarchy of desires; (2) the three problems the account of the extended ideal presented in Part I engendered; and (3) the ideal of autonomy.

1. Frankfurt's account fails to solve the problem of the endless hierarchy of desires (the regress problem)

Is volitional necessity able to fulfil the role Frankfurt gives it, namely providing the limit to the hierarchy of desires and therefore being essential to autonomy? *Do* love and care really stop the endless hierarchy of desires and provide a limit to my will? I summarise the objection that arises from the previous points in this chapter where I have raised the 'normative authority' question.

Frankfurt's account lacks scope for posing the evaluative, normative question: Is my love or my care *right* or *good*? The capacity to make such an evaluation is part of my autonomy. I care about my cares, and therefore I may question them. Frankfurt needs to make sense of that. This is possible only with reference to normative elements, for example to morality and principles. We *do* wish to evaluate our love and care and feel that we ought to (be able) to judge others concerning their love and care. As I have argued previously, this wish and feeling is part of what it is to be autonomous – self-direction includes self-evaluation. To direct yourself means to know what you want, and you can only achieve this sort of wholeheartedness if you know what you care for. (Frankfurt could agree with this.) But, as an autonomous being, you are also able to evaluate those deepest cares themselves. I do not believe Frankfurt can make sense of that radical self-evaluation without weakening the 'necessity' and 'authority' love imposes on our will, and abandoning his sharp distinction between ethical decisions about one's life, on the one hand, and one's love and care, on the other hand.

My objection that the normative dimension is lacking is based on the assumption that it is essential to maintain the normative, evaluative dimension as part of the ideal of autonomy. Some cares are morally good, others not. But why, Frankfurt could reply, do we need that as part of the ideal of autonomy? Why is normativity or normative authority needed? Volitional necessity offers a way of stopping the infinite regress without reference to something outside the agent and it is a metaphysically economical solution. Is it not a fact that people experience volitional necessity? My objection is that the mere fact that I love this or that does not make this love a *necessary* fact: I could love something/someone else. Why should my love be the love it is? Frankfurt could reply to this that this normative question does not matter for the question of autonomy, since my love is informing my choice without being my choice (it is given), and that this solves the infinite regress problem. Is this not ideal? My reply is that this is not ideal because it admits of loves and cares we might not want ourselves or other people to have. It is part of being autonomous to judge your own cares. But what if two autonomous persons have cares that result in conflict? There is a problem with conflict resolution if (the consequences of) my love conflicts with (the consequences of) your love. Frankfurt, however, could propose finding practical solutions for these matters, without having to refer to morality. This point is questionable: to solve a conflict between people who both care about different things one may need to refer to what is 'right' or 'fair' or 'good' – to some kind of common normative standard. But even if we accept it, and therefore conclude that Frankfurt's account is not flawed (if construed in the way I have done) then it remains true to say that his ideal fails to account for the fact that humans aware of 'volitional necessity' are not prevented from questioning their deepest attachments, being afraid of having the wrong attachments, etc.

This could be understood in two ways. Firstly, we could see it as a flaw in Frankfurt's account in so far as 'volitional necessity' is to stop the infinite regress of desires. If I question my love, then as a result of my reflection and evaluation I may produce another higher-order desire which destroys my wholeheartedness by generating struggle; there is no longer a limit to my will, and the infinite regress of higher-order desires re-emerges. In other words, the concept of 'volitional necessity' collapses. Secondly, however, even if we disregard this argument, and give credit to Frankfurt's ideal for acknowledging the 'necessary' nature of certain of our attachments, it remains true that Frankfurt is not able to make full sense of the normative and evaluative

aspect of our lives, which includes looking at the status of our loves and cares from a normative point of view. This does not mean, as Sartre would have it, that we can always still *choose* and *act* differently, but rather that although our deepest attachments (making us the person we are) impose themselves on us as a 'necessity', constraining our will, compelling us to take a certain course of life, we are able (precisely because we are autonomous beings) to question those attachments and evaluate them in the light of consequences, principles, reasons, values, etc., in other words, in a normative light. An ideal of autonomy which disregards this aspect of human being is therefore – although perhaps still *an* ideal of autonomy – *inferior* to one that does deal with the question of the normative authority of our desires.

Furthermore, by disregarding strong evaluation as an aspect of the constitution of a person, Frankfurt also fails to ask the question what can constrain, limit, assist, and guide this strong evaluation.

Note that it may be, of course, that often we do not engage in this sort of activity and do not even want to. I suggest that this is a kind of 'Bad Faith': we are deceiving ourselves about our capabilities as human beings if we *never* (want or dare to) question what or whom we love and care for.

The search for a constraint to the evaluation of love and care could lead us back to the extended ideal of autonomy, of course, an ideal which prompts reference to 'God' or 'good' as the ultimate limits and guides to our self-evaluation. But since we were in search of (1) a metaphysically economical ideal, (2) an ideal that does not make me dependent on something outside myself, and (3) an ideal that deals with the question whether we, as autonomous beings, can (still) choose evil, this is not to be pursued at this point. It is, therefore, necessary to continue the inquiry.

2. Frankfurt's account does not deal adequately with the three problems the extended ideal left

In relation to the problems with the extended ideal, Frankfurt's account leaves out one very important problem, namely Problem Three – or, at least does not provide a satisfactory solution of that problem.

In Frankfurt's account it is perfectly possible for a person to be aware of his evil love (say love for killing) and still do evil (still kill), since he might be a true lover of evil, and it is the necessity of his will to do so. But how can we know that something is evil and still choose it? If the answer is: 'no, this is not possible', then Frankfurt would need to show that a person is not able to know that he loves evil and still do evil. But

this is excluded in his concept of volitional necessity. If I 'cannot do otherwise' I am very much aware of this volitional necessity, I know it, I feel it. The constraint on my will, as I read Frankfurt, is not unconscious at all. The only way out would be to claim that the lover of evil is deceiving himself, since his 'real self', 'what he really is', is not evil. But then Frankfurt needs a criterion to distinguish 'real' from 'not real', in other words, something that can guide us to decide whether our love is the right love to have, whether what I care for really is 'me'. So Frankfurt's answer has to be: 'Yes, it is possible to know that something is evil and still choose it.' But then he needs to give an account of how this is possible; in other words, an adequate account of the grounds for choosing evil. Such an account is lacking in Frankfurt's work, unless one accepts 'what I really care for' and 'who I really am' as being able to provide such a ground.

3. Frankfurt's account cannot be construed as an ideal/the best possible ideal of autonomy

The following argument summarises my attempt to support a stronger claim against Frankfurt's account of autonomy as an ideal of the person, stronger than saying that it is not flawed but merely incomplete in the sense that it does not do justice to our desire to connect the ideal of autonomy with an ideal of the good. I believe more can be said. In making the former points I have repeatedly tried to argue that there is something about *autonomy* missing in Frankfurt's account. Let us recapitulate this argument.

The ideal of autonomy as self-direction necessarily excludes our making certain choices. By directing ourselves in one way, we always exclude other ways (at least temporarily and most of the time definitively – there may be no way back). This may be partly what Sartre means by his claim that we 'have' to choose; the problem with this claim, however, is that Sartre assumed that in making any choice I am not bound by *anything*. I have rejected that notion of radical choice. So now, with regard to the self-direction of our lives, we can ask the question what constrains (and therefore helps, guides) us in our choice. Firstly, as Frankfurt argues, there is what I love and what I care for. Secondly, there is, not unrelated to the first constraint, the social, cultural, historical, and ideological context I live in (Frankfurt acknowledges this). But thirdly, there is the fact that these former constraints do not *determine* me. I am able, after all – to a certain extent, certainly not always and absolutely, and in so far as I possess the full range of human capacities – to question what and who I care for and

love, evaluate my deepest attachments, etc. This ability to question and evaluate myself is itself part of the ideal of autonomy: I wish to have this ability of deep self-evaluation and exercise it. Then the question emerges: what can guide me in that activity? The answers to this question explored in the first part of this book have raised serious problems. Therefore we continue to look for an alternative. But it now appears clear that Frankfurt's work does not provide the resources to deal with this dimension of human reality and of autonomy. This renders his account not flawed as such, perhaps, but less complete *if it is construed as an account that is to inform a viable modern ideal of autonomy*. It says something about autonomy, but it does leave out an important aspect of autonomy.

Furthermore, if Frankfurt not only leaves out an important aspect of autonomy, but fails to deal with the problem of the endless hierarchy of desires and with Problem Three, I conclude that Frankfurt's account of volitional necessity is not a suitable alternative to the extended ideal developed in the first part of the book. As I said before, if this 'strong claim' is justified, this provides a good reason to continue to look for an alternative.

Frankfurt is not always descriptive; he does not hesitate to make a normative claim about love, since he writes that 'the reason we must not betray what we love is that we must not betray ourselves' (Frankfurt 1999: 174). The 'musts' in his sentence are indicative of the normative force Frankfurt wants to give to love. In the context of the ideal of autonomy, speaking about whether or not we betray ourselves is very relevant: autonomy has to do with identity, with what is 'me' and what is 'not-me'. But it is striking that in making his claim Frankfurt assumes that we can change neither ourselves nor what we love.

Key to my strongest objection to Frankfurt's account is, then, the claim that a person's identity is not fixed. But what does this mean? It cannot mean that my identity is 'floating' in the sense of being wholly dependent upon social context, culture, etc., as social determinism would have it, since then there is no sense in which I can call myself autonomous. This relates to Problem Two, namely the question whether if I depend on something outside me I am still autonomous. If what I am is completely socially *determined* then I am not autonomous. But the answer that there is no fixed identity to a person need not mean complete determination by something outside me. There is another possibility, namely what is (or could be) part of the modern ideal of autonomy: self-direction through strong evaluation, understood as the

evaluation of who I am and what I care for. Then my identity is not entirely 'fixed' or 'stable' but it is still (to some extent) 'up to me'. I admit that such an activity, in which I am the object as well as the subject, may challenge the limits of what we are capable of as humans. It may be very difficult and near to impossible. But *if* it is possible, there is no reason why it should not be part of a rich and complete concept of autonomy and an ideal of autonomy to aspire to.

6.5. Conclusion

Out of the objections we can distil a weak and a strong claim. The *strong claim* is that Frankfurt's account is not flawed as such, but if construed as an ideal of autonomy, (1) it fails to account for an aspect of autonomy, in particular 'strong evaluation' of the person's deepest cares and attachments; (2) it fails to put a limit to the hierarchy of desires; and (3) it does not deal (well) with Problem Three of the extended ideal. In other words, it may do as a descriptive moral psychology (an incomplete one), but cannot function as an ideal of autonomy. The *weak claim* is that there is nothing wrong with Frankfurt's ideal of the person as such, but that since it is unconnected to an ideal of the good, it may be unacceptable to many people as an ideal on its own. (Applied to the example of the committed Nazi, this is the view that even if the committed Nazi is an ideal *autonomous* person according to the Frankfurtian account, he is not an ideal person, since in this view being an ideal person includes living up to a moral ideal apart from being autonomous.) Furthermore, if Frankfurt's ideal of autonomy neglects the role of the normative in our self-understanding and our understanding of others, it is at least in that sense incomplete.

On the basis of the arguments made above, I think I have made a good case for the stronger claim. This stronger claim can itself be split into a stronger and strongest claim. Firstly, Frankfurt's account is not flawed as such but if construed as an ideal of *autonomy* it fails (for the reasons mentioned above). (This is the stronger claim.) Secondly, his account is flawed as such. If Frankfurt's concept of volitional necessity does not work, that is, fails to provide a limit to the infinite regress problem, then Frankfurt's account is internally incomplete. I admire the fact that Frankfurt tries to reconcile our wish to be free and autonomous with our experience that sometimes we cannot do otherwise. However, his attempt to do so, by introducing the concept of volitional necessity, fails. In a sense, Frankfurt went too far in one direction, that of necessity and constraint, neglecting our capacity for

deep self-evaluation. In another sense, he went too far in the other direction, that of freedom, since precisely by understanding our deepest cares as *given* (constraints) and as something that's a very *personal* matter, he gives free play to those currents in 'liberalism' that are all too happy with the view that morality is a private matter and that nobody has business with what I love or care for, therefore *de facto* arguing for more freedom in the sense of 'caring for what you want' without any objective or shared ways of judgement and evaluation. If viewed thus, Frankfurt does not contribute directly to a better understanding of autonomy nor provide us with better objections against what he calls 'the ideal of a society that maximises the freedom of its members to do what they want' (Frankfurt 1999: 156) than the idea that we ought to be discriminatory in our desires. To connect autonomy with being discriminatory in what you care about, in other words, with deep self-evaluation and judgement, and to inquire into the basis for such an evaluation and judgement (what can guide me?) seems to me a more fruitful approach, even if this may mean a costly metaphysics. However, I hope that I can find in Kant a better way of reconciling autonomy and volitional constraint, and a better way of dealing with the problems related to the extended ideal developed in Part I. (This is the strongest claim.)

Frankfurt's view of love and care and the volitional necessity they can produce is admirable in its efforts to account for those human experiences that give us the feeling that our will is constrained, that we 'cannot do otherwise' – and this not as the result of some pathological compulsion or addiction but because we have certain cares and loves which make us the person we are and make us unable to act otherwise. I have argued, however, that this psychology – however accurate as a description of these particular experiences – is unable, as it stands, to say something about this other human possibility: the ability to subject our deepest attachments to normative and moral evaluation. To the extent that this ability is constitutive of our essence as an autonomous person, Frankfurt's account fails. And because of this 'normative gap' it also fails to provide an adequate limit to the potentially endless series of higher-order desires; I will always be able, in principle, to question my deepest attachments and form a higher-order desire about them (I like my love, I don't like my love, I want to change my attachment, etc. – regardless of the question whether I actually *can*). Furthermore, Frankfurt neglects the question whether a person, caring for something or somebody, may be morally *wrong* in his caring, and he therefore neglects, accordingly, the related question (related to

Problem Three) whether if I know that something is evil I can still genuinely *care* for it or even *love* it.

In the next chapter, I will look at Kant's ideal of autonomy and consider what he can offer to deal with the problems identified for the extended ideal developed in the first part of this book. I will argue that his account, unlike the others I have treated so far, is particularly relevant to attempting to resolve Problem Three.

Introduction to the Next Chapters: Two Kantian Ideals Of Autonomy

In the following chapters, I will consider two Kantian ideals of autonomy. By doing this, it is not primarily my aim to provide a comprehensive view of Kant's view of autonomy or of his moral philosophy. Rather, I shall deal with the questions 'To what extent is this view Kant's own view?' or 'What did Kant really say?' only in so far as this assists my search for an alternative to the extended ideal of autonomy which is able to deal with its problems, drawing on Kant and Kantian material.

The first view I will consider is Hill's Kantian ideal of autonomy, since it may provide a metaphysically economical solution to the problem of infinite regress and therefore avoid Problem One of the extended ideal (that of a costly metaphysics).

The second view I will consider is based on my own construction of the ideal of autonomy to be found in Kant's *Groundwork* and *Religion within the Limits of Reason Alone*, aided by Allison's interpretation. I will show that Kant was well aware of the problems I have identified in my account of the 'extended' ideal of autonomy and made a serious attempt to deal with them.

7
Hill's Ideal of Autonomy

7.1. Introduction

Frankfurt's ideal of autonomy seemed at first sight a suitable alternative to the extended ideal of the person, but I have shown its limitations and therefore the need to continue the search for a further approach. Hill's ideal of autonomy is worth considering for this role, since, as I will show, it presents us with a metaphysically economical ideal of autonomy close to the heart of what I have argued we aspire to if we wish to be autonomous persons.

Firstly, I will reconstruct Hill's argument, showing what he tries to achieve and how far he succeeds. Secondly, I will evaluate this achievement in terms of Hill's own aims. Thirdly, I will discuss possible objections to Hill's idea of choice and deliberation. Finally, I will argue that in the end Hill's ideal of autonomy is unable to fulfil the role we may wish to give it as an alternative to the extended ideal of autonomy developed in the first part of this book, which therefore remains – despite its problems – the best account we have so far. (I shall, in the next chapter, examine further challenges to it by considering Kant's own ideas.)

7.2. Hill's Kantian ideal of autonomy

The development of Hill's argument consists of two 'movements'. One is a negative one, constructing the boundaries of the concept of autonomy 'from without', defining the ideal by excluding others, that is, by saying what it is *not*. The other movement is definition 'from within', a positive construction starting off with defining the meaning of autonomy and relating it to various elements, building up a network of related meanings meant to express the Kantian ideal of autonomy.

First, I will present the two 'movements', then show the merits of Hill's ideal, considering the extent to which he reaches his aims and solves the problems of the extended ideal of autonomy.

7.2.1. What the Kantian ideal of autonomy is not, according to Hill

Hill (1991) argues that Kantian autonomy has to be distinguished from the following ideals: (1) the ideal of autonomy as a psychological capacity and the highest stage in the moral development of a person; (2) the ideal of Sartrean autonomy; and (3) the ideal of autonomy as a right.

1. The ideal of autonomy as a psychological capacity and the highest stage in the moral development of a person

Firstly, it is undeniable that autonomy defined as a psychological capacity and the highest stage of moral development could be influential in the construction of autonomy as an *ideal*. If, as Piaget[43] saw it, 'heteronomy' means that external rules are followed, rules laid down by other people, and 'autonomy', by contrast, means that the rules guiding the person's behaviour are the outcome of his own free decision, then surely this second stage of development is ideal, that is, something we ought to aspire to. A similar move from the descriptive to the normative, from the 'is' to the 'ought' occurs in relation to Kohlberg's theory.[44] Pre-conventional and conventional levels are characterised by rule-following because of the consequences, because of the physical power of the one who rules, or because the person conforms to social expectations. By contrast, the highest level of 'autonomy' is the one in which the person defines moral values and principles by a decision of conscience, 'appealing to logical comprehensiveness, universality, and consistency', as Duska and Whelan put it (Duska and Whelan 1977: 47). Surely this last (stage in the last) level is seen by Kohlberg as ideal? Consider the moral character traits defined as morally mature and as belonging to the autonomous person: reflectiveness, self-control, independence of judgement, commitment to general principles, emotional independence from others, special propensity for abstract thinking, and exceptionally critical attitudes towards current social norms. The reasoning behind developing this ideal of autonomy as moral maturity may be summarised as follows:

> If a person spends his whole life doing what he has been told to do by authority, merely because of fear of authority (stage one), or

> because it will bring him pleasure (stage two), or because it is expected by the group (stage three), or because that is the law (stage four), he has never really made moral decisions which are his *own* moral decisions. [...] One must be one's own person, so to speak, in order to mature fully. One must develop one's own principles of judgement and action. (Duska and Whelan 1977: 69)

Is this not part of what it is to be autonomous? In his essay 'The Kantian Conception of Autonomy' (1989), Hill observes that the 'idea of autonomy as a variable psychological trait serves for some as both a descriptive category and a normative ideal for moral agents' and that this ideal (at least in Kohlberg's version, as summarised by Duska and Whelan)

> includes not only reflectiveness, self-control, and independence of judgement, but also commitment to general principles, apart from hope of reward or fear of punishment; and it requires making moral decisions from this loyalty to abstract principle rather than from compassion for particular persons. Autonomous agents, on this view, make moral decisions from an impartial perspective, detached from the special feelings that stem from their particular personal relationships. (Hill 1989: 92)

In *Autonomy and Self-Respect* Hill argues that on this view autonomy is defined as the 'capacity and disposition to make choices in a rational manner; and this means choosing in the absence of certain particular attitudes and inner obstacles, such as blind acceptance of tradition and authority, neurotic compulsions, and the like' (Hill 1991: 31). Hill gives examples of nonautonomous persons and behaviour such as children accepting authority without questioning, adolescent rebellion against authority, traditionalism, and compulsive gambling. According to Hill, this does not mean that we have to act independently of all causes and desires, but independently of those causes and desires which interfere with rational choice. Hill explicitly refers to the ideals associated with this conception of autonomy:

> The ideals naturally associated with this conception of autonomy are the development of rational capacities in education, the overcoming of unconscious psychological disabilities through psychotherapy, and the *use* of one's rational capacities in making important choices. (Hill 1991: 31)

Hill admits that we can recognise in the ideal of autonomy as moral maturity (and its associated ideals) several ideas Kant would have applauded. However, he argues that there are 'crucial differences' between these psychological conceptions of autonomy and Kantian autonomy (Hill 1991: 31). Note that if there are indeed these crucial differences, it does not follow that psychological conceptions cannot constitute an ideal of autonomy. It is not because the ideal of autonomy as psychological maturity is not *Kantian* that it cannot be *an* or even the most cogent and compelling ideal of the autonomous person. Are there reasons independent of whether it is Kantian or not that could support not considering it as an adequate ideal of autonomy? Let us first look at Hill's arguments as to why it is not a Kantian ideal.

Firstly, Kant does not see autonomy as an 'empirically discernible trait, attributed in various degrees to people on the basis of what they are observed to say and do in various circumstances' but rather as something 'attributed on a priori grounds to all rational wills' (Hill 1989: 93). According to Kant, autonomy is not a special achievement but rather a universal condition of moral agency.

Secondly, Hill argues that several features of psychologically autonomous persons are not essential for Kantian autonomy. According to Hill, 'emotional independence from others, special propensity for abstract thinking and exceptionally critical attitudes towards current social norms' are not required. Furthermore,

> even acting from internalised moral principles, contrary to social norms and without concern for reward and punishment, would not guarantee the possession of Kantian autonomy for that autonomy requires acknowledging the principles not only as 'self-imposed,' in some sense, but also as unconditional requirements of *reason*. (Hill 1989: 93)

Hill takes the Kantian view to be that even those who are knowingly immoral and those who are loyal to individuals rather than impartial principles 'still have wills with the property of autonomy, though of course they fail to express their autonomy by living up to the commitments it entails' (Hill 1989: 93).

So although undoubtedly many of us may *de facto* view the character traits related to autonomy as psychological maturity as those of an ideal autonomous person, Hill's critique shows that to call this ideal 'Kantian' is very problematic. But this, by itself, does not make it inadequate as *an* ideal of the autonomous person, as I have said.

There is, however, a good reason why we need not here consider this ideal further. It is my aim in this part of my book to look for an ideal of autonomy that could replace the extended ideal developed in Part I on the grounds that it is able to deal with the problems I have identified. The ideal of autonomy as psychological maturity is unable to perform this task. Both in the form of a psychological theory based on empirical evidence and in the form of vaguely defined character traits that *might* be seen by *some* people as expressing the ideal of autonomy, it is certainly metaphysically economical (it supposes only empirical evidence (as a theory) and 'common' sense (as a 'common' view)) but it is not equipped and not remotely able to tackle the questions of how autonomy and dependency are related, and whether autonomy includes the freedom to choose evil.

Certainly, the theory may be able to give an 'answer' to these questions (based on empirical evidence, for example: the results you get when you ask people what they think about autonomy), but, by itself, it is unable to provide *arguments* for those answers. It lacks the philosophical depth and resources to provide good reasons why we have to understand autonomy in this way.

Nevertheless, I have discussed Hill's arguments concerning psychological maturity since, firstly, it is a metaphysically economical account of autonomy and since, secondly, it is part of Hill's larger argument concerning *Kantian* autonomy – assisting us, therefore, in trying to understand Hill's construction of a Kantian ideal of autonomy. By its metaphysical economy and its partial resemblances to Kant (through elements such as universality and principles) it carried the promise of being an adequate Kantian ideal of autonomy, but we know now that it is neither Kantian nor adequate to the purpose.

2. The ideal of Sartrean autonomy

I have discussed the Sartrean ideal of autonomy in an earlier chapter. How could this ideal be confused with Kant's? Let us look at the reasons Hill cites as to why it might have gained popularity among philosophers: 'Suppose one is attracted by Kant's idea that one is morally bound by nothing but what one imposes on oneself and also by his denial of determinism regarding human choices, but one cannot accept Kant's noumenal/phenomenal distinction, his moral rigorism, or his belief in principles of conduct which are essentially rational for everyone. A natural result would be acceptance of what I shall call Sartrean autonomy' (Hill 1991: 30). What is left is a theory which says that people are autonomous if, firstly, their choices are not causally determined (compare with 'having

a free will') and, secondly, 'people are morally and rationally free to do as they choose in that there are no objective values, only self-imposed commitments. No general moral principles follow from the contention that persons are autonomous in this sense' (Hill 1991: 31).

I agree with Hill that this ideal is obviously different from Kant's. Kant held that autonomy is a property of the will, and such a will requires, according to Kant, acting on principles. The key point is that I am committed to principles not because I desire to follow them or because they are expected to lead to something I desire, but 'simply by virtue of being rational. The principles are self-imposed insofar as they stem from one's rational nature rather than from fear of punishment, desire for approval, blind acceptance of tradition, animal instinct, and so on. To have *autonomy* of the will is to be committed to principles in this way and to be able and disposed to follow them' (Hill 1991: 29).

3. The ideal of autonomy as a right

The ideal of autonomy as a right means, according to Hill, that one has the right '(a) to make one's own decisions about matters deeply affecting one's life, (b) without certain sorts of interference by others, (c) provided certain conditions obtain. The right presupposes a background of other moral rights and legal rights within a just system, which define an area of permissible conduct' (Hill 1991: 32).

Hill discusses this ideal at some length, but with regard to its relation to Kant he says that 'insofar as the idea has roots in Kant, it stems more from his principles of liberty and respect for persons than his metaphysical doctrines. Though other rights have been associated with autonomy, the right I have in mind is a moral right against individuals' (Hill 1991: 31–2). So we can conclude that this ideal does not have a direct connection with Kantian autonomy at all. However, Hill discusses it since it is part of *his own* Kantian ideal of autonomy. Let us turn to this ideal now.

7.2.2. What the Kantian ideal of autonomy is, according to Hill

Hill's 'Kantian' ideal of autonomy consists of at least five elements or aspects: volitional unity, understanding of one's own values, deep deliberation, free choice of what to value, and the development of one's capacities for rational self-control.

1. Volitional unity

Hill starts with the definition of autonomy itself. Looking at what he takes to be the core meaning of the word 'autonomy' itself, self-governance, he suggests that 'people are not self-governing, in a sense, when

their responses to problems are blind, dictated by neurotic impulses of which they are unaware, shaped by prejudices at odds with the noble sentiments they think are moving them. When we make decisions like this we are divided against ourselves' (Hill 1991: 50).

Note the parallel with Frankfurt's ideal of wholeheartedness. Both ideals require the will to be united. This requirement is different from the claim that we ought to have a *strong* will. In 'Weakness of Will' Hill offers a good argument for the claim that it is not enough to have a strong will to be an ideal person; it *may* be a necessary but certainly not a sufficient requirement. Strength of will may be a virtue, but having this virtue is conditional upon that person having also other virtues: 'Strength of will would be an ideal to strive for along with charity, justice, etc., but it might be worthless by itself' (Hill 1991: 137). I maintain that strength of will is perfectly compatible with immoral attitudes and beliefs; consider for example a strong-willed Nazi (as considered in several instances earlier).

2. Understanding your own values

Hill argues that compassion can be the guiding value in relation to making and reaching decisions. Of course it has to be genuine compassion and not 'a self-deceptive mask for concern for reputation' (Hill 1991: 51). So agents need to be aware of the relevant features of a moral problem and have an effective understanding of their real values. Autonomy as an ideal is, according to Hill 'neutral in disputes about which is more important, compassion or respect for rights. What it tells us is merely that we should try to face moral decisions with integrity and self-awareness' (Hill 1991: 51).

3. 'Deep deliberation'

Hill says something about our capacity for 'deep deliberation' in his essay 'Pains and Projects' (Hill 1991). (I have already discussed Taylor's related concept of strong evaluation (Section 1.4 and Section 3.2.).) What does Hill mean by deep deliberation? If the capacity to engage in deep deliberation is to be part of the ideal of the autonomous person, it does not mean that we are supposed *constantly* to question everything. Hill answers that we can't, but 'as Kant, Aristotle, and others remind us, we can review in turn the main ends we take for granted in ordinary deliberation and ask whether it would be rational to make them our ends apart from how pursuing them affects our resources and other ends' (Hill 1991: 175). So we could say that ideally, according to Hill, we ought sometimes to engage in this process of deep deliberation

as opposed to ordinary deliberation. Ordinary deliberation is, as I interpret Hill, deliberating about *means* rather than ends, or about, say, choosing between my desire for chocolate and my desire for beer. Deep deliberation, on the contrary, is not about my love of chocolate, not about 'highly variable first order values' (Hill 1991: 177), but about my final preferences and ends: 'Deep deliberation,' therefore, 'searches for the basic grounds for choosing' (Hill 1991: 178). This distinction between first-order values, on the one hand, and what we may call 'higher-order values' is certainly very useful. But what are 'final ends'?

Hill seeks to specify what one must consider as a final end by excluding what is, according to him, *not* to be regarded as a final end. Firstly, he excludes what he calls 'harmless desires, like wanting the home team to win' (Hill 1991: 181). Nor do more serious concerns, like attachment to individuals and career, 'which are central to the sort of person I now am and want to be' (Hill 1991: 181) survive scrutiny: I still have to determine whether I must 'count it as a reason in deliberation' (Hill 1991: 181). Note the difference with Frankfurt, since the latter concerns would count for him as *cares* and to these cares Frankfurt gives a much higher status (that of volitional constraints).

Furthermore, Hill rejects pleasure and pain as the 'necessary common denominator of rational choice' (Hill 1991: 181). He grounds his argument by considering the contrary view that it is impossible not to care about pleasure, that we are to some extent drawn by it. Now this may be true, but, according to Hill, this 'being drawn' 'has no more necessary rational status than the *desires* we have just considered. Perhaps I cannot help caring but in deliberation the fact that I care need not be counted as a reason even *tending* to justify the corresponding choice' (Hill 1991: 182). This comment provides a good basis for criticising Frankfurt's account, and echoes my objections to Frankfurt made earlier (Section 6.4.). The fact that I cannot help caring about something cannot be considered a final end or a reason in the active review and examination *of* one's inclinations, desires, and cares.

Hill admits that one may find that one cannot discount certain factors, and gives the example (just as Frankfurt does) of Martin Luther's remark 'Here I stand, I can do no other.' But as Hill sees it, rather than powerlessness this expresses 'sustained commitment' (Hill 1991: 183). As I read Hill, this suggests that Luther's remark is not about a 'volitional necessity' but rather expresses commitment which arises out of deep deliberation.

Let us look again at Hill's 'ideal of rational (deep) deliberation':

> first, in choosing ends, one must critically scrutinise one's actual and potential ends, in the light of one's best available information, asking whether one can rest content not only with the outcome of the selection but also with oneself as the person who made it; then, without assuming any substantive principles about what is necessarily a reason for choosing, one must *simply decide* what ends *at this moment* one can best justify to oneself; after this, one must rely on familiar instrumental principles for the choice of the best means to one's ends (Hill 1991: 183).

We already know what Hill does *not* consider as final ends; from this formulation of the ideal of deep deliberation we can infer Hill's account of what final ends *are*: the ends one at this moment can best justify to oneself. This 'oneself' is important in relation to autonomy. Hill's formulation of the character of deep deliberation says something about autonomy as the capacity to make my *own* choices. One needs to be content with *oneself* as the person who *made* the choice and it is about justification *to oneself* and one's *own* ends. In other words, it is *me* who chooses: this is self-direction, self-governance. (Note that Hill's account does not spell out on what basis we can 'justify' ends to ourselves. He only excludes 'substantive principles about what is necessarily a reason for choosing,' saying that we must '*simply decide*' (Hill 1991: 183). But what is it to 'simply decide' then? See my objection later in this chapter (Section 7.5.).)

Hill's remarks about Kant help us in developing further a Kantian version of the extended ideal of autonomy. For example, he mentions the Kantian idea that 'in deep deliberation, rational agents necessarily want to *respect themselves* as agents' (Hill 1991: 186). In general, Hill's book rightly connects autonomy and self-respect. Part of what it is to choose appropriately, for ideal autonomous persons, is that their choices stand up to 'the most thorough critical scrutiny of and by themselves' (Hill 1991: 186). The ideal is that 'each person should choose in such a way that he can maintain his self-respect over time' (Hill 1991: 186). This is important for my argument in favour of a strong link between the ideal of autonomy and moral constraint; this requirement for self-respect to be sustainable 'over time' means that we *cannot* just care about anything:

> One might, just conceivably, maintain self-respect for a time while choosing to pick blades of grass instead of more challenging human

> pursuits, but it is less likely that one could do so, in deep reflection, for long. (Hill 1991: 187)

Hill writes that 'the standard of what to value about the life of others […] should *prima facie* be *their own choices*, unless those choices cannot withstand their own critical scrutiny (which we are rarely in a position to know) and unless, on reflection, their choices are ones that we cannot aid without losing our own self-respect' (Hill 1991: 188).

4. Free choice of what to value

Hill says more about the free choice of agents about what to value in relation to the ideal of autonomy as a right I mentioned earlier. Autonomy as a right does not mean the absolute freedom to do what you want. Autonomy as a right 'is limited, for example, by principles of justice, noninjury, contract, and responsibility to others' and we can take decisions that deeply affect our life 'so long as they are consistent with other basic moral principles, including recognition of comparable liberties for others' (Hill 1991: 48). In saying this, Hill seems to hold a 'neutral' notion of autonomy which then needs supplementation by a moral framework or defines such a framework within which the individual is then free to do what he wants. A similar 'neutral' notion has been proposed by John Rawls.[45] The idea is that 'a theory of autonomy, following Kant, Rawls, and others, would first define principles for moral institutions and personal interactions, leaving each person, within these constraints, the freedom to choose and pursue whatever ends they will' (Hill 1991: 42). In conclusion, 'the right of autonomy allows people some room to make their own choices; it does not dictate what those choices should be' (Hill 1991: 49).

This conclusion tempts Hill to suggest the following link between this ideal of autonomy as a right (which is also part of his ideal) and Kant's. What Hill means by autonomy is that the agent 'can, within a wide area of life, choose what to value and what not to value without contravening any fixed, objective, pre-set order of values in the world. As Kant, Sartre, and others have maintained, an autonomous person is a "creator of values"' (Hill 1991: 97).

Note that Hill's ideal of autonomy is both different from but also comparable to Frankfurt's, since he maintains that *within a range of morally permissible choice* 'we may choose to value some things, and to disvalue others' (Hill 1991: 97). It is different, since it allows the realm of morality to limit my choice. It is comparable, since it still makes a sharp distinction between morality and free choice of what to value, a

distinction which roughly corresponds with Frankfurt's distinction between morality and what I care for.

5. Development of one's capacities for rational self-control

Hill's discussion of the capacity for rational self-control tells us more about what could be part of a Kantian ideal. He writes, for example, that the modified Kantian principle he proposes 'commends the development of one's capacities for rational self-control, not simply for the results, but because this is a natural expression of valuing for their own sake one's capacities as a rational autonomous agent' (Hill 1991: 100). This means, for example, that suicide is opposed to this principle, since it treats life as a rational agent as instrumental to a maximum pleasure/pain balance. Hill refers to the latter view as a 'Consumer Perspective', which asks the question 'What will I get?' rather than 'What can I make' of my life? (Hill 1991: 100). Indeed, 'getting what you want' is a (utilitarian) principle often operative in people's decision-making. But it is obviously not necessarily part of a coherent ideal of autonomy; Hill defends a Kantian view as opposed to a utilitarian view.

7.3. Merits of Hill's ideal: the extent to which he achieves his aims and solves the problems of the extended ideal of autonomy

Hill effectively expresses many modern concerns and wishes that often crystallise around the concept of autonomy. Many of us hold it as ideal to have a unified will (knowing what we want), an adequate understanding of one's own values (knowing what we care for), to engage in deep deliberation (being able and wishing to question our own aims), to have free choice about what to value (we want and have the right to make our own decisions about our life and what sort of persons we want to be), and to undertake the development of one's capacities for rational self-control (we value and use our capacities we have as rational agents).

The inclusion of the idea of deep deliberation in particular makes it possible to see Hill's ideal of autonomy as being a real step forward in comparison with Frankfurt's account. If autonomy requires that the choice persons make stands up to 'the most thorough critical scrutiny of and by themselves' (Hill 1991: 186), there is a fair chance that it also 'stands' as a limit to the endless regress of desires. If I have subjected my choice to 'the most thorough critical scrutiny', why would I still

(be able to) form higher-order desires? I deeply deliberated about my choice, so therefore if *this* is my choice then *this* is what I want. Full stop.

(I will later on in this chapter (Section 7.6.) give my reasons why Hill's account is not able to provide this 'full stop' but at this stage of my argument it is important to see that Hill's concept of deep deliberation could be seen as a step forward in comparison with Frankfurt. Frankfurt's concept of volitional necessity didn't work, or so I argued. Here we are provided with a concept here that promises to provide what is needed.)

Hill also deals implicitly with the three problems identified for the extended ideal of autonomy.

1. Metaphysical economy

Hill's ideal of autonomy is certainly metaphysically economical. Without having to resort to the view of autonomy as the highest stage of moral development, his account does not bear the cost of an elaborate metaphysics. He successfully incorporates what he might call 'the best of Kant', the ideal (and wish) to be 'morally bound by nothing but what one imposes on oneself' and the denial of determinism regarding human choice. In contrast to Sartre, this does not lead him to reject Kantian essentialism (I mean Kant's view that we are essentially rational beings) in favour of the 'groundless choice' view. Still, for Hill to agree with Kant that principles are self-imposed insofar as they stem from one's rational nature does not commit him to embrace an extravagant metaphysics. His discussion of Kantian and Sartrean autonomy (Hill 1991: 30–1) suggests that Hill is aware of the problem of metaphysical economy (Problem One) and that he takes the view that a Kantian theory of autonomy need not be embedded in a metaphysical framework which would be unacceptable to most philosophers today. The Platonic and Augustinian extension to the modern ideal of autonomy suggested in the first part of this book certainly involves such a framework. According to Hill, 'Kant's theory suggests less encumbered ideals of autonomy which continue to have a wide appeal' (Hill 1991: 30).

2. No dependency on something outside me

Hill's ideal of autonomy need not assume dependency on something 'outside' the person. We have seen that Hill holds the view that the person is a 'creator of values' and 'can, within a wide area of life, choose what to value and what not to value without contravening any

fixed, objective, present order of values in the world' (Hill 1991: 97). According to Hill, it seems, the only external limit to *decisions* and *actions* based upon our choice of what to value, or our creation of values, is of a *practical* nature. If we live in society, our decisions are limited by 'principles of justice, noninjury, contract, and responsibility to others' and they need to be consistent with 'other basic moral principles, including recognition of comparable liberties for others' (Hill 1991: 48). In other words, to make our living together possible, there need to be these constraints. In this way, the criterion of 'no dependency on something outside me' is satisfied *so far as practically possible* given that we live in a society. But are these limits 'just' practical? What do I mean by 'practical'? The limits Hill mentions have a moral character (as well). Therefore, Hill needs to question whether his claim that there aren't any contravening fixed, objective values in the world is compatible with the limits he proposes. To put it more precisely: Hill needs an account of why and how we (as individuals and as a society) are bound by principles of justice, noninjury, contract, and responsibility to others and by other basic moral principles such as the recognition of liberties for others if those moral principles are not based on a fixed, objective order of values. (We shall see when considering objections to Hill's view (Sections 7.4 and 7.5.) that the problem I note here is symptomatic of related difficulties in other aspects of Hill's view.)

2. Can I still choose to do evil if I know something is evil?

In his essay 'The Kantian Conception of Autonomy', Hill argues against the view he ascribes to Kant that the will 'cannot freely choose between acting from inclination and acting from moral principle because the will, as practical reason, is directed exclusively toward what is rational and moral' (Hill 1989: 95) on the grounds that if this were true, all 'immoral' acts would prove to be not even willed by the agent. If, when autonomous, our will were always 'on the side of the angels' (Hill 1989: 95), then if we are not autonomous we would not be under moral obligation. We would be 'carried away' without being responsible for an action since it is unwilled. To avoid these consequences, we have to accept the view that the will can still choose between acting from inclination and acting from moral principles and that it is part of what it is to be autonomous to be able to make this choice.

Is this a sound argument? I will present my objection to it below. Furthermore, the question whether I can choose between *inclination* and acting from moral principles is different from the question

whether I can choose *evil*. Hill's account lacks the resources to deal with this question. But, before discussing objections related to the question whether Hill is able to provide an answer to the three problems with the extended view of autonomy, I would like first to ask whether Hill achieves his own aims.

7.4. Why Hill fails to achieve his own aims: Is Hill's ideal Kantian?

If it was Hill's aim to develop a *Kantian* conception of autonomy, he only partly succeeds.

Firstly, *can* we, according to Kant, still choose between acting from inclination and acting from moral principles? In the next chapter I will cast some serious doubts on this claim. But whatever Kant's answer is, it will be shown that Kant at least *deals* with it, and with the related problem of choosing evil, whereas Hill does not. (See further discussion below: why Hill fails to deal with Problem Three.)

Secondly, Hill suggests a far too great proximity between Kant's ideal of autonomy and the ideal of autonomy as a right. The notion of autonomy as a right does not *in itself* demand morality, as the Kantian notion does, but rather allows for moral constraints operating, so to speak, at the boundary of one's own decisions about one's life. Morality may be constraining, but constrains only 'from without', not 'from within' as part of what being autonomous itself means. Hill's words that 'the right of autonomy allows people some room to make their own choices; it does not dictate what those choices should be' (Hill 1991: 49) reveal a conception of autonomy which refuses a role for morality to constrain the choice of the individual 'from within'. Essentially, within such a conception the wants, desires, and preferences of individuals are taken as 'given', they are not meant to be constrained 'before' they are made subject to general rules of 'justice' and 'permissible conduct'. The spirit of Kantian autonomy, I believe, is however rather to bear upon the individual's desires and wants 'right from the start'. There is not first self-directing choice and then constraint, but the self-direction is *itself* constrained by universal principles of morality, or, rather, *is* what it is to be moral*, is* what it is to be governing oneself by rational universal principles.

Hill modifies Kantian autonomy in such a way that it is compatible with the idea of choosing what you want. In saying that an agent is autonomous, Hill means in part that the agent (as I have cited before) 'can, within a wide area of life, choose what to value and what not to

value without contravening any fixed, objective, pre-set order of values in the world. As Kant, Sartre, and others have maintained, an autonomous person is a "creator of values"' (Hill 1991: 97). Surely Kant would not have agreed with this. In a footnote Hill admits that Kant and Sartre differ on the sense in which we create values: for Sartre this creation is free from objective rational constraints, whereas for Kant it is (obviously) not. But this difference is fundamental, I think, and in fact means that we cannot attribute the label 'creating one's own values' to the ideal of Kantian autonomy. As I read Kant, there is self-governance but not creation of your own values. In following universal principles, we do not create these principles and neither have I created my rational nature on the basis of which I exercise this autonomy.

Hill *seems* to defend a Kantian view as opposed to a utilitarian view. But although he rejects 'getting what you want' he leaves much room for 'doing what you want' or in any case 'doing with your life what you want' and 'choosing/caring for what you want'. Although such a view does not automatically allow in utilitarianism, it at least encourages it. Since Hill does not conceptualise constraints to 'what I care for' and 'what I do with my life', he leaves room for the person to decide upon these questions in a utilitarian way. I do not say that this is good or bad, but, rather, that it contradicts Hill's effort to provide a *Kantian* account of autonomy.

Third, Hill's concept of deep deliberation misses the universal aspect of choice, and, to that extent, it misses a key Kantian focus. I will say something more about this below (Section 7.5., objection 6).

However, although this section has helped to point to difficulties with Hill's view already, it is not my principal aim to criticise Hill's view because it fails to capture key aspects of Kant's thinking, but rather for reasons intrinsic to his view that make it unacceptable as an alternative to the extended ideal of the person. Firstly, I will point to the intrinsic problems with Hill's view of choice and deliberation. Secondly, I argue why his view, taken as an alternative to the extended ideal, fails to solve the problems of that ideal.

7.5. Objections to Hill's idea of choice and deliberation

Firstly, the example I gave of the strong-willed Nazi suggests that we may want to link Hill's notion of unity of will with the demands of morality. If this is done, the ideal of autonomy as volitional unity seems to correspond with my notion of 'inner autonomy'. However, in Hill's version it leaves much more room for all kinds of decisions and

behaviour. In particular, the difference is that Hill's notion of autonomy as unity of will only includes unity of will as opposed to a very limited amount of situations of 'self-division', namely situations when my response to problems is dictated by neurotic impulses:

> People are not self-governing, in a sense, when their responses to problems are blind, dictated by neurotic impulses of which they are unaware, shaped by prejudices at odds with the noble sentiments they think are moving them. When we make decisions like this we are divided against ourselves. (Hill 1991: 50)

This definition of self-division suggests that as long as I see the real situation, I'm fine. With the word 'blind' Hill does not refer to the opposite of moral vision or awareness of 'reality' (in a Platonic sense) but to the opposite of considering the real situation in the sense of having an undistorted perception: 'Ideally autonomous, or self-governing, moral agents would respond to the real facts of the situation they face, not to a perception distorted by morally irrelevant needs and prejudices' (Hill 1991: 51). I agree with Hill that this should be part of an ideal of autonomy, but I believe it is not enough. It is conceivable that my perception is not distorted, that I'm not blind in Hill's sense, that I'm not neurotic, etc. but *still* I may be not autonomous in the sense that I'm not able to make a moral judgement. Next to seeing the morally relevant facts there needs to be a *moral judgement* – but can this judgement be genuinely moral if it is not guided by rational principles? On the basis of Hill's formulation of unity of will/self-division, we can infer that Hill wants Kantian autonomy but without constraints to the will other than seeing the relevant facts.

Secondly, in relation to achieving an adequate understanding of your own values as part of Hill's ideal of autonomy I noted that Hill's ideal is 'neutral in disputes about which is more important, compassion or respect for rights' (Hill 1991: 51). But is such a neutral notion satisfactory? Hill does not consider the question of whether a person's real values are also the *right* or *good* values *tout court*. As long as they are real (in the sense that these are the values the person *really* has), all is well. But what is the source of our values? And how are we to decide which values we make our own? Is this possible without an ultimate point of reference (as opposed to just another value)? Don't we need some guidance about what is more important?

Thirdly, Hill's idea of deep deliberation involves a distinction between first-order values, on the one hand, and what we may call

'higher-order values'. Deep deliberation is not about my love of chocolate, that is, not about 'highly variable first order values,' (Hill 1991: 177) but about my final preferences and ends: 'Deep deliberation', therefore, 'searches for the basic grounds for choosing' (Hill 1991: 178). But does Hill really get down to the 'basic grounds for choosing'? That would involve the question: Isn't there a level higher than the one of 'higher-order values'? Isn't there a need for an ultimate point of reference to decide which higher-order values are to be preferred? Hill neither asks nor answers these questions.

Fourthly, there are further problems with Hill's formulation of deep deliberation. Consider again this formulation of the concept (already quoted above):

> first, in choosing ends, one must critically scrutinise one's actual and potential ends, in the light of one's best available information, asking whether one can rest content not only with the outcome of the selection but also with oneself as the person who made it; then, without assuming any substantive principles about what is necessarily a reason for choosing, one must *simply decide* what ends *at this moment* one can best justify to oneself; after this, one must rely on familiar instrumental principles for the choice of the best means to one's ends (Hill 1991: 183).

There are problems with this formulation at every stage. Firstly, it is not clear on what basis one is able to scrutinise one's ends. Secondly, it is not clear how one can *simply decide* about ends without principles or another basis. Thirdly, in the choice of our means too we might want to apply moral principles; not all means are morally good.

Hill's view of choice is superior to the Sartrean one since it does not reject *all* grounds for choosing, but clearly this formulation neglects the *ultimate* grounds for choosing. Any account of autonomy and/or deep deliberation needs to say something about this. Even if the formulation would suggest that our choice is ultimately unconstrained, taking the existentialist position, the question about the ultimate grounds for choosing cannot be avoided. Hill's view of choice sits unhappily between the constraints of Kantian morality and the liberal attachment to (nearly absolute) freedom.

Fifthly, the requirement attached to deep deliberation that one should choose in such a way that one can maintain self-respect 'over time' certainly constrains our care, but seems to me very minimal. We want a strong(er) link between the ideal of autonomy and moral

constraint. What if a serial killer claims to maintain self-respect by killing? Of course, to this objection Hill could reply that we ought to act within a legal framework of basic rights and justice. But what if I decide to spend my life as an alcoholic, regularly taking a legal drug? What if I decide for myself – in deep deliberation as Hill understands it – that this 'maintains my self-respect over time'? If I decide on this life, do I exhibit the achievement of the modern ideal? I have already argued (see my discussion of Frankfurt) why we want to question our own (higher-order) values and why this is part of the modern ideal of autonomy. Hill's criterion does not enable us to carry out *this* deep deliberation, that is, deep deliberation properly understood as including questioning my deepest attachments. Self-respect (alone) does not seem to me a sufficient guide to question what I really care for and – ultimately – what I am. (See my earlier discussion of autonomy and identity.[46]) It is conceivable that to be able to question my identity I have to 'bracket' my self-respect at that point, since if I have too much respect for what I am I won't be able to change myself. I may have to ask myself the question: Is what I respect about myself *really* to be respected? Does this or that aspect of myself *really* deserve my respect? And it is plain that an answer to *this* question cannot rely on the criterion of self-respect; we need something else that can guide us.

Finally, the question whether a certain aspect of myself deserves *my* respect can be extended to the question whether this aspect deserves the respect of *others* as well. There is no (fundamental) difference between these questions, since there is always a universal aspect included in the person's *own* choice. My argument for this claim is as follows. The question whether an aspect of myself deserves the respect of others prompts the question whether and on what basis others can judge me and I can judge others. Therefore, I first discuss what Hill says in relation to judging others; then I proceed from this discussion to defend my claim about the universal aspect in a person's choice, using Kant's concept of self-respect.

As noted earlier, Hill says about judging the life of others that

> the standard of what to value about the life of others [...] should *prima facie* be *their own choices*, unless those choices cannot withstand their own critical scrutiny (which we are rarely in a position to know) and unless, on reflection, their choices are ones that we cannot aid without losing our own self-respect (Hill 1991: 188).

Firstly, it may be objected: Why does our being able to examine other people's choices matter? There may be good reasons why this matters. In my discussion of Frankfurt, I have argued that autonomy includes (rationally) judging *my* values, choices, and actions, and that, therefore, to the extent that autonomy matters to me, what others do matters too. Since in judging myself I ask what could be willed by all, I do judge others too. I think that this is part of what Sartre meant when he said that in choosing for me I choose for all. Another argument is that because, when judging and justifying my actions (for) myself, I reason, I can – in principle at least – communicate these reasons to others. I can justify myself to others. And why should I do so? Because in practice we do see that people want to see reasons, especially since and if my actions affect others. In particular, in a liberal society I have to justify *any* form of interference with another's liberty. The key thought seems to me to be that if it is part of autonomy to judge myself I should also respect other people's right to do so. And this entails that it matters what I do and what others do. So at least I want to be able to judge if other people's values may interfere with mine.

Secondly, however, Hill's standard does not allow judgement of others at all. If we cannot know if people's choices can withstand their own critical scrutiny, we are not able, on the basis of this criterion, to judge the values and lives of others at all. The last point about aid and self-respect also does not work as a criterion, since it does not provide guidance as to what aid to what choices *ought* to threaten our self-respect.

If we want to construct a viable ideal of autonomy, of which deep deliberation is a part, we need to make sure that the *universal* aspect of the person's *own* choice is not kept out of sight. Certainly, deep deliberation as part of self-governance is something that *I alone* can do, and it is or should be finally *my* choice what I do with my life. However, this cannot mean that the way I make this choice is or ought to be unconstrained by anything apart from my desires, preferences, or cares. The Kantian idea of self-respect is that I respect myself *as a human being*, and therefore I have to ask myself what could be a universal principle for *all*. This universal element within the personal process of deliberation and self-guidance should not be neglected. That people often do not take this universal position in 'deep' deliberation may be a description of how some people live, but if they do not take this position then it is not genuinely *deep* deliberation and it is not really the exercise of *autonomy*, let alone can it be the *ideal* of the autonomous person. This universal and Kantian aspect of the ideal does not imply,

as some proponents of liberty may suspect, that I tell you what to do; ideally, we all have the liberty to make our own choices. Rather, since in making my own choice I ask the question whether what I choose can be chosen by all, this deliberation is a safeguard against authoritarianism and tyranny, since these forms of life are not universalisable. By not incorporating the moral aspect to autonomy, whether Kantian (as being considered here) or Platonic, an ideal of autonomy tends to leave too much room for 'doing what you want' which turns that ideal of 'autonomy' into a certain 'ideal' of freedom perhaps, but not in the end an ideal of autonomy.

7.6. Why Hill fails to solve Problem Three of the extended ideal

As I have shown above, if Hill's alternative were unproblematic it would solve the first two problems of the extended ideal. As it stands it is metaphysically economical, and to 'simply decide' about final ends does not require, in Hill's view, reference to something outside myself. But I have shown that this *is* problematic, and the above discussion suggests that to solve his problems he should include in his view of choice reference to something outside 'me', a reference to something that is not metaphysically economical. However, to spell out what this 'something' could be I would have to refer to the extended ideal developed in Part I (and this is not appropriate since we're looking for an alternative) or take my next step (which I postpone for my conclusion and my next chapter). Therefore, at this point, I will let this rest and argue how Hill fails to deal adequately with Problem Three, since little has been said about this yet.

Consider again his argument as I presented it above. Hill's conclusion that we can still choose between acting from inclination and acting from moral principles and that it is part of what it is to be autonomous to be able to make this choice remains problematic. He fails to ask and answer the question: On what basis can we make *that* choice?

Furthermore, he suggests that from his conclusion it follows that autonomy is not strongly connected with morality. According to Hill, it follows that a rational agent is not committed to the categorical imperative, nor is it a necessary rational requirement 'that we try to maximise desire-satisfaction, the balance of pleasure/pain, or any other substantive value that, as human being or as individuals, we happen to care about' (Hill 1989: 100). The latter claim is arguably true, and

provides a good argument against Frankfurt's view of what we care about as a necessity that may constrain our will without reference to anything further. But it is not clear that the first claim – that a rational agent is not committed to the categorical imperative – (1) follows from the conclusion above that we are still free to choose and (2) can be claimed to be Kant's (or at least a Kantian) view. With regard to (1), consider the following objection. I may still be free to choose between inclination and duty, but *as a rational agent*, I am committed to choosing the latter. It may be, as Hill himself writes, that people 'fail to express their autonomy by living up to the commitments it entails' (Hill 1989: 93), but from this it does not follow that their autonomy is *only* related to the choice between 'good' and 'evil' and not *also* to the principle(s) they are committed to as rational agents. With regard to (2), namely how this claim is related to Kant's view, in other words, if this is an adequate interpretation of Kant, I refer to further discussion below (the next chapter on Kant).

To conclude: even if we follow Hill's argument up to the point of the conclusion that we are still free to choose, I hope to have shown that (1) this conclusion is not an end-point and is still very problematic (Problem Three is not answered) and (2) Hill's following argument, namely that therefore autonomy and morality are not strongly connected, is not conclusive.

7.7. Conclusion

At first sight, Hill's attempt to construct a Kantian ideal of autonomy answers Problem One of the extended ideal of autonomy (lack of metaphysical economy). He also seems to deal with Problem Two (dependency on something outside the person) by embracing the 'autonomy as a right' view. However, (1) I have given reasons why Hill fails in part to achieve his own aim to construct a *Kantian* ideal, (2) I have pointed to the intrinsic problems with his view of choice and deliberation, and (3) I have argued that to solve these problems reference to something outside 'me' is needed, a solution which is not metaphysically economical and therefore fails to answer Problem One. Finally, in trying to deal with Problem Two by embracing the 'autonomy as a right' view, Hill runs into the difficulty of not being able to account for Problem Three. The result of his 'neutral' ideal of autonomy is that Hill remains confronted with the following problem. To put it in Kantian terms, if it is part of the ideal of autonomy that a person still has the possibility not to choose out of duty, as Hill argues, the question needs

to be answered how that person can make a choice between duty and inclination. A neutral ideal cannot provide any criteria, reasons, or guidance which may help the person in choosing. If his ideals of unity of self and deep deliberation are to make sense, they have to be embedded in a theory of the ideal of autonomy which solves or avoids the internal problems caused by the requirement of 'neutrality' and accounts for all three problems.

In the following chapter I present my own construction of Kant's ideal of autonomy based on the *Groundwork of the Metaphysic*[47] *of Morals* (1785), *Religion within the Limits of Reason Alone* (1793), and the *Metaphysics of Morals* (1797),[48] aided by Allison's interpretation (1990). I will show that Kant's theory deals with the other problems in a way unmatched by Frankfurt's and Hill's account of the modern ideal of the autonomous person. In particular, we find in Kant a serious attempt to answer Problem Three (with his theory of *radical evil*).

8
The Ideal of the Person in Kant's *Groundwork*

8.1. Introduction

The (primary) aim of Kant's *Groundwork*[49] is neither the development of a certain notion of autonomy nor the defence of a certain ideal of the person. Rather, it is 'to seek out and establish *the supreme principle of morality*' (Gr 392). Moreover, as we will see, Kant does not write about autonomy as a capacity or property of the *person* but rather of the *will*. In spite of these particular points, however, I claim that we can detect in the *Groundwork* an ideal of the person and an ideal of the *autonomous* person.

This ideal, if understood in terms of acting according to reason, promises to be metaphysically economical. Furthermore, Kant's very notion of an autonomous person appears to exclude dependency by definition, since, as we will see, it means that the will gives the moral law to itself. Finally, Kant – in contrast with many of the previous authors – makes a serious attempt to deal with Problem Three (the question of whether autonomy and the choice of doing evil are compatible).

Firstly, I will make explicit the Kantian ideal of autonomy on the basis of my analysis of the *Groundwork* and discuss the relations (similarities and differences) between this ideal and the extended ideal of the autonomous person as developed in my book so far. I will also discuss links with the Kantian ideal of the person in the *Doctrine of Virtue* as interpreted by Allison. Secondly, I will show how Kant deals with Problem Three. An investigation of the *Wille/Willkür* distinction and the problem of radical evil will take us to the *Groundwork* again but also to *Religion within the Limits of Reason Alone*. Again, to the extent that I go beyond the *Groundwork*, my discussion will be aided by Allison's interpretation.

8.2. The ideal person according to Kant (based principally on the *Groundwork*)

8.2.1. Principles and reasons

In attempting to establish the supreme principle of morality, Kant assumes in the *Groundwork* that morality is a matter of following principles. It is not my aim to question this assumption, but to show what follows from it for the ideal of the person. Kant writes:

> Moderation in affections and passions, self-control, and sober reflection are not only good in many respects: they may even seem to constitute part of the *inner* worth of a person. Yet […] without the principles of a good will they may become exceedingly bad; […]. (Gr 394)

This shows that Kant holds moderation, self-control, and sober reflection to be the characteristics or virtues of an ideal person, but not unconditionally. If a person is to act, it is *principled* action alone that makes the action good. Kant reinforces this point in relation to the Christian command to love your neighbour. Love may be a virtue, but again not unconditionally. Kant argues that the command only makes sense if the love in question does not depend on or arise from inclination (alone):

> For love out of inclination cannot be commanded; but kindness done from duty – although no inclination impels us, and even although natural and unconquerable disinclination stands in our way – is *practical*, and not *pathological*, love residing in the will and not in the propensions of feeling, in principles of action and not of melting compassion; and it is this practical love alone which can be an object of command. (Gr 399)

Note that this contrast between 'practical' and 'pathological' love is not made by Frankfurt. But this is not the only difference between Kant and Frankfurt. Their accounts differ fundamentally. Frankfurt's volitional account is unconnected with the idea of duty, principled action, practical reason, etc. Love may reside in the will, and even constrain the will, but for Frankfurt questions of morality, duty, etc. are distinct. For Kant, on the contrary, love residing in the will *is* 'kindness done from duty', practical love. Kant would argue that much of what Frankfurt means by love is 'pathological' love and therefore not residing in the will but in feelings.

Furthermore, note also that Kant's concept of practical love is very different from Plato's idea of love of the good. The former is love (of somebody) commanded by duty, the latter is love of the good (itself), uncommanded. It does not follow from this, however, that therefore the Kantian idea of morality is radically incompatible with the Platonic extended ideal of the autonomous person. My arguments below will show that there are important points of connection between the two.

A first attempt to reconcile them could be to give the Kantian notion of the (highest) moral principle a place within the Platonic extended ideal.[50] Kant's categorical imperative (Gr 402) can be seen as the supreme principle guiding moral action. Although what Kant calls 'ordinary human reason' does not actually conceive this principle thus abstractly, 'it does always have it actually before its eyes and does use it as a norm of judgement'; it is a 'compass' to distinguish good and evil (Gr 403–4).

This claim provides me with an element that could be accommodated within the version of the extended ideal of the autonomous person developed so far. If the good is the 'magnetic pole', then we need a compass to find it, and the categorical imperative suits this role if it is able to distinguish between good and evil. I write 'if' since arguably there are severe problems with this Kantian claim. But that does not influence the main argument of my book. The only thing I'm arguing is that *if* we want to entertain the ideal of the autonomous person, then we need to take on board some idea of a 'magnetic pole' and, therefore, some idea of a 'compass'. Kant's categorical imperative is a good candidate to provide that, since although we specified already the 'magnetic pole' (the good) we haven't specified yet 'the compass' and the precise role of human reason. Kant's account of morality is very helpful here.

Let's assume that Kant's compass works, that it is able to distinguish between good and evil. Then the problem is that although we might have this compass, this way of distinguishing between good and evil, this way to find the good, and therefore in consequence *know* the good, we may still not act accordingly. Plato would deny this possibility. If I really know the good, how can I do evil? Augustine struggled with this, and so did Kant. I will return to Kant's answer to this problem in my discussion of his *Wille/Willkür* distinction and his concept of radical evil. In the *Groundwork*, Kant sees the problem as a tension between inclination and duty, and ultimately as a tension between happiness and morality. Happiness he defines as the total satisfaction of all one's needs and inclinations (Gr 405);

morality, analogously, we could define as the total satisfaction of (the demands of) duty.

The contrast between happiness and morality provides us with a useful tool for a comparison between the Kantian notion of morality, on the one hand, and the view that you can do what you want as long as you do not harm others, on the other hand. We met this latter notion of morality and autonomy in the discussion of Hill's ideal of autonomy and in the discussion of Frankfurt, the view that morality is a matter of the public sphere and within the private sphere I may do what I want if this or that makes me happy. Kant would not restrict (questions of) morality to the public sphere. Rather, it seems to me that, on his view, questions of morality are relevant to both spheres, and that we need to make a strict distinction between questions of morality and questions of happiness.

But to return to the main problem here: the inclinations, according to Kant, provide 'a powerful counterweight to all the commands of duty' and from this arises 'a *natural dialectic* – that is, a disposition to quibble with these strict laws of duty, to throw doubt on their validity or at least on their purity and strictness, and to make them, where possible, more adapted to our wishes and inclinations' (Gr 405).

Firstly, note that this analysis of human morality is more refined than Augustine's. Whereas Augustine seems to see only two possibilities – either indulge in lust and sin, or behave according to the eternal laws – Kant allows for a dialectic and in this way accounts for the possibility that humans doubt the principles of morality themselves. To put it simply: if the Augustinian steals a pear, (he reproaches himself afterwards since) he knows he has failed to do good. If someone who thinks within a Kantian framework steals a pear, he will afterwards start doubting whether it is really so bad to steal at all, and whether there isn't an exception to the principle for stealing a pear. This would not be an ideal person according to Kant, of course, and therefore this tells us something more about the Kantian ideal, the absence or minimisation of such a dialectic.

Secondly, we see that Kant is a very good observer of modern morality, and he here anticipates an explanation for how we arrive at the Nietzschean and Sartrean view that there are no (*a priori*) moral principles. It can be argued that once we start wondering whether there aren't some exceptions to the principle, we start doubting that principle itself. And once we start doubting the validity of a principle, we actually do not only doubt its content but its *existence* as such. Given the nature of a moral principle, it is impossible to genuinely believe

that there exists a principle X and at the same time to claim that there are exceptions to X, that we do not always have to follow X. This I take to be Kant's point when he writes that to adapt principles to our wishes and inclinations is to 'pervert their very foundations and destroy their whole dignity' (Gr 405).

This point leads Kant from the study of 'ordinary human reason' to moral philosophy: what is then the foundation of the principles? And how do principles function? We, however, are interested in the ideal of the person that emerges in the discussion. So far, we can conclude that Kant's ideal of the person is one of a *reasonable* person following *principles* and therefore doing good, being in control, being moderate, etc. Kant's moral philosophy favours rational concepts, since 'a mixed moral philosophy, compounded of impulsions from feeling and inclination and at the same time of rational concepts […] can guide us only by mere accident to the good, but very often also to the evil' (Gr 411). This, together with what has been said before, provides us with the ideal of the person who is guided by reason to the good, using the categorical imperative as a 'compass'.

I have remarked already that Kant is writing about the *will* rather than the person. Given this perspective, how does he define the relationship between will and reason in the *Groundwork*? And what does this tell us about the ideal of the person? I suggested earlier that Kant's 'will' is not the same as Frankfurt's 'will'. In fact, in the *Groundwork* Kant identifies will with practical reason itself. He defines will as 'a power to choose *only that* which reason independently of inclination recognises to be practically necessary, that is, to be good' (Gr 412). How can we read this? We could infer that this is Kant's ideal of the person: a person who has this will and exercises this will. Kant recognises that in reality humans are not always ideal in this sense. He employs a contrast between objective necessity and subjective contingency. If reason itself is not sufficient to determine the will, if the will is exposed to certain impulsions 'which do not always harmonise with the objective ones', 'then actions which are recognised to be objectively necessary are subjectively contingent' (Gr 412–13). This Kant labels 'necessitation' – in contrast to necessity – to refer to the relation of objective laws to 'the will of a rational being' as one in which the will, 'although it is determined by principles of reason, does not necessarily follow these principles in virtue of its own nature' (Gr 413). So although there are commands (of reason), although there is an 'ought', we do not always follow them and the reason for this lies in our own nature as humans. If we were divine, Kant argues, we would not need

imperatives or 'oughts': 'for the *divine* will, and in general for the *holy* will, there are no imperatives: '*I ought*' is here out of place, because '*I will*' is already of itself necessarily in harmony with the law' (Gr 414). From this we can infer that Kant holds it to be ideal for a person to have a divine or at least holy will (I will not discuss the precise difference). However, since he doesn't think that this is reachable for humans, we can hardly conclude that this is his ideal of the (human) person. How can we be supposed to follow an unreachable ideal? There are two possible answers to this problem.

One answer is to say that Kant still thinks it to be an ideal we should aspire to, even if we cannot reach it, and therefore it *is* his ideal of the person. So, although holiness is unobtainable, we have a duty to strive after it, and this (striving) is what it is to be virtuous. In *The Metaphysics of Morals* Kant writes about man's duty to increase his moral perfection. This perfection consists 'in the *purity* (*puritas moralis*) of one's disposition to duty, namely in the law being by itself alone the incentive, even without the admixture of aims derived from sensibility, and in actions being done not only in conformity with duty but also *from duty*. Here the command is "be holy"' (Me 446).[51] Taken on its own, this is a problematic claim, since it is difficult to make sense of the pursuit of an unattainable ideal. Allison argues: 'Since one cannot sincerely adopt any maxim without acting according to it, one cannot adopt a maxim of holiness without striving with all one's power to realise this ideal. […] But holiness, as we have seen, is unattainable by finite beings. […] Accordingly, it turns out that we are morally required to pursue an unattainable goal; and this seems absurd' (Allison 1990: 171). In reality, however, we see that people *do* (want to) pursue unattainable goals: compare the Christian moral requirements. The solution to this problem, then, is to see holiness as an *ideal*, admitting that, as Allison does, 'the requirement to orient one's life in a certain direction, namely, toward an ideal, is one that can never, in principle be completable' (Allison 1990: 178). So our duty is to strive after holiness, not to attain it. This is the conclusion Kant reached, since on the same page he qualifies the claim I quoted above: 'It is man's duty to *strive* for this perfection, but not to *reach* it (in this life), and his compliance with this duty can, accordingly, consist only in continual progress' (Me 446). Elsewhere in the *Metaphysics* Kant stresses this idea of progress when confirming the paradoxical character of virtue as an unattainable ideal: 'Virtue is always *in progress* because, considered *objectively*, it is an ideal and unattainable, while yet constant approximation to it is a duty' (Me 409). In *Religion within the Limits of Reason Alone* Kant writes

that the man who adopts 'doing one's duty, merely for duty's sake' as his maxim is not yet holy by reason of this fact alone, since there may be a great gap between the maxim and the deed, but still 'he is upon the road of endless progress towards holiness' (Re 42).

There are further problems with this view, in particular with the claim that we have a duty to act from duty. Ross has argued that this involves Kant in an infinite regress. If in relation to every duty we also have a further duty to do it from duty, then that principle must apply to that further duty *ad infinitum* (Ross 1930: 5). There may be more 'internal' problems with this view. Our question, however, is whether the ideal of holiness needs to be part of the (extended) ideal of the autonomous person. A possible argument for a positive answer to this question could be the following. Since guidance by the good is a necessary condition for the ideal of autonomy to be coherent, the ideal of autonomy includes that we try to reach the good. Therefore, *if* we want to be autonomous, we are – in Kantian terms – 'morally required to do all in our power to realise the Highest Good' (Allison 1990: 172).

Another answer to the problem of holiness as an unreachable ideal, an answer not necessarily completely incompatible with the first, is to say that, apart from this unreachable ideal, Kant has a *second-best* ideal more suited to humans, which is informed by the first ideal and therefore not completely incompatible with the first, but different from it since it lowers the demands. This second ideal is that there *is* an imperative – we don't do good 'automatically' so to speak – and that we *follow* this imperative. So the essential difference is that 'I will' is not necessarily in harmony with the principle, so therefore it becomes an 'I ought', and if we then follow the law, in other words act from duty, then we are 'ideal' persons. I will say more about Kant's second-best ideal below in my discussion of self-control.

Note that this second-best ideal is distinct from 'doing what you want' in the sense of following your inclinations. It concerns 'the good', not 'the pleasant': the 'practically good [...] is distinguished from the *pleasant* as that which influences the will, not as a principle of reason valid for every one, but solely through the medium of sensation by purely subjective causes valid only for the senses of this person or that' (Gr 413). Kant's ideal of the person is distinct from the ideal of the happy person. Kant argues that even if we tried to attain happiness, 'the concept of happiness is so indeterminate a concept that although every man wants to attain happiness, he can never say definitely and in unison with himself what it really is that he wants and wills' (Gr 418). This is the ultimate reason why the pursuit of happiness as

such cannot be an ideal of the person. A person would sometimes do one thing, sometimes another, and in every case have no certainty that happiness would result. A person aspiring to such an 'ideal' has 'no principle by which he is able to decide with complete certainty what will make him truly happy, since for this he would require omniscience' (Gr 418). With Kant we can conclude that 'happiness is an Ideal, not of reason, but of imagination' (Gr 418).

8.2.2. Autonomy

It has been noted already that for Kant autonomy is not a property of the person but of the will. Nevertheless, if we make the assumption that an autonomous person is a person with an autonomous will, we can proceed in discussing Kant's concept of autonomy in relation to the ideal of the autonomous *person*.

Kant, as noted, equates the will with practical reason itself, the will which follows the principles of universal law. However, this needs at least two qualifications. Firstly, we can only understand this claim as saying that *in so far as* we are rational beings, we follow the principles of universal law. The former discussion has shown Kant's way of accounting for the fact that human wills do not necessarily follow the universal law. In so far as we're sensible beings, we are inclined to follow our inclinations.[52] Secondly, the claim that in so far as we're rational we *follow* the universal law, our wills are *necessarily determined* by reason, etc., suggests a problem for the question of autonomy. If we understand autonomy along the lines suggested so far in this book, it is not entirely clear why this Kantian ideal (following the law, being determined by the law) is an ideal of the autonomous person.

Firstly, am I autonomous in the inner sense? On the one hand, the answer may be 'yes'. If I follow the universal law, if I'm determined by it, an inner order and harmony emerges. My acts, desires, will are all in accordance with the universal law. On the other hand, however, this creates the problem that in so far as I'm not a fully rational being, not ideal, not divine, not holy, there is a 'natural dialectic', to use Kant's term, and the inner order and harmony disappear. So there is a question with regard to whether we can achieve this ideal of autonomy. But in general, we can conclude that the Kantian ideal is an ideal of 'inner' autonomy, reachable or not.

But is it also an ideal of outer autonomy? Am I still independent if I'm determined by the universal law? One way to go would be to say that it doesn't make me dependent on other *people*, and that it even gives me more independence from 'opinion' since I have an independ-

ent, universal standard by which to judge. This is true, I think, and I argued in this way earlier for the extended ideal of the autonomous person. However, the Kantian ideal requires more work, since we need to make sense of terms such as 'necessity' and 'determination' – as opposed to *guidance* which is the term I have used up till now to qualify the relation between 'good' and 'me'. How much outer autonomy do I have if I'm – in the ideal case – determined by universal law? Kant's answer to this would be the following: (in the ideal case) we can not only consider ourselves as followers of the universal law, but at the same time as its *authors,* its *makers*. Kant's ideal of the person includes 'the Idea *of the will of every rational being as a will which makes universal law*' (Gr 431). This Kant calls the principle of autonomy.

How can we understand this claim? For Kant, autonomy is not only a property of the will, it is also a principle in itself, very similar (Kant suggests identical) to the categorical imperative: '*Act only on that maxim through which you can at the same time will that it should become a universal law*' (Gr 421). This imperative is implied in the formula or principle of autonomy quoted above: 'By this principle [the principle of autonomy] all maxims are repudiated which cannot accord with the will's own enactment of universal law. The will is therefore not merely subject to the law, but is so subject that it must be considered as also *making the law* for itself and precisely on this account as first of all subject to the law of which it can regard itself as the author' (Gr 431). *Because* I can regard myself as the maker, the author of the law, I am subject to it.

As suggested above, this notion of autonomy fits with the extended ideal of the autonomous person in the following way. To be able to be autonomous I need to be guided by the good. To be able to be guided by the good, I need a 'compass' to find the good. This compass could be Kant's categorical imperative. The categorical imperative gets me to the good by requiring me to direct my actions on the basis of universal principles of reason. The nature of this relationship is one of necessity and determination, but only on condition that I can regard myself at the same time as the maker of the principles. Then I'm not merely subject to the law or to interests which are not my own; the law springs at the same time from my own will and is universal. Such a will then is not merely guided by the good, but *is* good. According to Kant, the categorical imperative and the formula of autonomy provide at the same time the formula 'for an absolutely good will' (Gr 437). We can infer that at the same time it provides the formula for a good person. The idea of guidance by the good, then, leads us to the idea of the good person as part of the ideal of the autonomous person. Furthermore,

Kant's idea of what it would be if all persons were good is comprised in the notion of a world of rational beings, a *mundus intelligibilis* as a kingdom of ends made possible 'through the making of their own laws by all persons as its members' (Gr 438). On the basis of the previous discussion this could be regarded at the the same time as a world of autonomous persons.

8.2.3. Good will and the good

I will now further compare the view of Kant's ideal of autonomy at which I have so far arrived with the ideal of 'doing what you want' and with the extended ideal of autonomy as developed in this book.

As I argued in Part I of this book, the danger in seeing autonomy as a property of the person, rather than a property of the will, is in my view that autonomy may be falsely understood as meaning 'doing what you want'. Therefore, we need to be clear that if we – in contrast to Kant – want to construe autonomy as a property of the person, we transfer the Kantian constraints on the will to the person. We could say that the Kantian version of the extended ideal of the autonomous person includes the following definition: an autonomous person is a person who has an autonomous will and is (therefore) guided by the categorical imperative. As with 'good' or 'God' so we have here principles that guide us, and since we're concerned with the *ultimate* – not just any good but *the* good and not just any god but *God* – this means in this case that autonomy entails guidance by the *ultimate* principle, which is, according to Kant, the categorical imperative.

Note that in our lives we may be guided by many things: people, reasons, values, feelings, etc. The idea of an ultimate guiding source, however, is, apart from guiding us directly, also meant to bring order in this 'world of guidance'. Certain sources of guidance are more worth following, and if two guides conflict there needs to be an ultimate guiding source to decide between them. I do not claim that there *is* such an ultimate guiding source; rather I argue that *if* we want to uphold the ideal of the autonomous person, we need such a guiding source.

Kant's view is, I believe, compatible with the Platonic version of the extended ideal of autonomy. It has been suggested before that although 'the good' may be the 'magnetic pole', we need a compass to find out about it, and Kantian universal reason (in the form of principles, and in particular the ultimate principle of the categorical imperative) seems to be a suitable candidate for such a compass. In other words, if the Platonic 'shortcut' to the good is not possible, there is a road through universal reason. According to Kant, good is a property of the will, but

our will is not always good, and our actions are not always good either. Let us take Kant's example of a man taking pleasure in helping others. Kant insists that only if the man acts 'without any inclination for the sake of duty alone; then for the first time his action has its genuine moral worth' (Gr 398). I interpret this claim as meaning that our will is or becomes good only if we act from duty. I understand Kant as arguing that if we act from duty, if we follow universal reason, we arrive at 'the good', and this good is not something 'out there' but becomes a property of our will, and indeed (here I possibly differ from Kant) of the person.

However, this way of relating the Platonic version of the extended ideal of autonomy to the Kantian ideal of autonomy attempted so far is only one way of approaching the problem. Instead of starting off from the extended Platonic ideal and then searching for a role for Kantian notions, we may want to approach the matter the other way round. In the first paragraphs of the *Groundwork* we find the idea that the only thing that is unconditionally good is a good will:

> It is impossible to conceive anything at all in the world, or even out of it, which can be taken as good without qualification, except a *good will*. […] A good will is […] good in itself. (Gr 393–4)

If we understand this notion of a good will as meaning a will governed by the moral law, a will directed by the principle of obedience to the requirements of the categorical imperative, then there is no difference between the pursuit of the good and the obedience to principle(s). For Kant, then, acting from principle is a good in itself. Whereas my former attempts understood principles as a compass which directs us to the good, a pathway to the good, this interpretation suggests that for Kant, principles, and in particular the categorical imperative, are *the* good in themselves. A Kantian ideal of autonomy by self-guidance would then not mean guidance by principles to the good, but rather guidance by principles directly, since they *are* good.

One difference between the Platonic and the Kantian ideal of the autonomous person is that the former suggests that – ideally – we are attracted by, and indeed love, the good, whereas the latter conceives of morality in terms of a law which commands rather than attracts me. We have not the capacity unfailingly to love the good, but the capacity to submit to the dictates of practical reason. The aspect of authority prevails over the aspect of love. The Christian version of the extended ideal, then, combines both aspects. The Christian god is a source of love as well as a source of (moral) authority.

8.2.4. Why Kant's ideal of autonomy is not morally 'neutral'

The Kantian ideal of the autonomous person can be expressed – similarly to Plato's – as an ideal of harmony. The following quotation summarises Kant's position and shows clearly that the ideal of the autonomous person if conceived in Kantian terms is anything but a morally 'neutral' ideal of autonomy but expresses what it is to be a good person and what we ought to aspire to:

> Thus morality lies in the relation of actions to the autonomy of the will – that is, to a possible making of universal laws by means of its maxims. An action which is compatible with the autonomy of the will is permitted; one which does not harmonise with it is forbidden. A will whose maxims necessarily accord with the laws of autonomy is a holy, or absolutely good will. The dependence of a will not absolutely good on the principle of autonomy (that is, moral necessitation) is obligation. Obligation can thus have no reference to a holy being. The objective necessity to act from obligation is called duty. (Gr 439)

Kant holds that 'the dignity of man consists precisely in his capacity to make universal law, although only on condition of being himself also subject to the law he makes' (Gr 440). This may or may not be true, but in any case it is a clear expression of an ideal of the person, and, given the condition Kant mentions, of the autonomous person. Kant makes the principle of autonomy the supreme principle of morality itself: 'the principle of autonomy is "Never to choose except in such a way that in the same volition the maxims of your choice are also present as universal law"' (Gr 440). It could be argued that if autonomy is a dominant ideal today, it may have little to do with this strong connection between autonomy and morality as construed by Kant, but rather with the understanding of autonomy as 'making my own laws' without reference to the categorical imperative. Kant, on the contrary, suggests that you need both parts to be able to call it autonomy, and indeed morality.

To clarify Kantian autonomy it is instructive to look at his contrasting notion of heteronomy:

> If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own making of universal law – if therefore in going beyond itself it seeks this law in the character of any of its objects – the result is always heteronomy. In that case the will does not give itself the law, but the object does so in virtue of

> its relation to the will. This relation, whether based on inclination or on rational ideas, can give rise only to hypothetical imperatives: 'I ought to do something because I will something else'. As against this, the moral, and therefore categorical, imperative, says: 'I ought to will thus or thus, although I have not willed something else'. For example, the first says: 'I ought not to lie if I want to maintain my reputation'; while the second says: 'I ought not to lie even if so doing were to bring me not the slightest disgrace.' The second imperative must therefore abstract from all objects to this extent – they should be without any influence at all on the will so that practical reason (the will) may not merely administer an alien interest but may simply manifest its own sovereign authority as the supreme maker of law. (Gr 441)

In contrast to Frankfurt, Kant identifies the will with practical reason itself. The volitional is linked with the reasonable. Autonomy means that I, as a rational being, 'go beyond myself' in the sense that I think about which maxims can be universalised, and *therefore* in fact I 'become myself'[53] since in this way I realise myself as a rational being. Heteronomy, therefore, means to alienate yourself from yourself. Heteronomy of the will means that 'the will does not give itself the law, but alien impulsion does so through the medium of the subject's own nature as tuned for its reception' (Gr 444).

Note the difference between this kind of 'reception' and the 'being taken over' in the Dionysian 'ideal' of the person: the former understands 'alien' as 'not in accordance with universal law', whereas the latter understands 'alien' as 'alien to the self'. The Kantian heteronomous person 'receives', but this reception is rather common for a rational person in control, to do something not because it's required by universal law but because of other reasons. And even if I do something out of inclination (as opposed to reasons, *any* reasons) I may be still in control, deciding I want to follow my inclination. The Dionysian rapture, however, means total loss of control and of self. The Kantian heteronomous person is not necessarily out of control of his self at all. If I act in a certain way not because it's required by universal law but because of other reasons, I'm still acting according to reasons, which is not at all being out of control. If my interest and motivation not to lie is not the maxim 'I ought not to lie' but my desire to maintain my reputation, this has nothing to do with being in control or not but rather with 'my will seeking the law that is to determine it anywhere else than in the fitness of its maxims for its own making of universal law'.

This shows more than anything else that the Kantian ideal of autonomy is not in the first place an ideal of *self-control* or even self-direction, and therefore different from, say, Berlin's expression of the modern ideal of autonomy. Rather, Kant's conception of autonomy is embedded in a theory of morality which expresses an ideal of the person as one whose motivation to follow principles lies in the principles themselves rather than in anything else. It may be, of course, that in order to exercise one's autonomy of the will self-control is a necessary condition. Furthermore, self-control may be a Kantian virtue in relation to not following (certain) inclinations. But as such, self-control is not the main feature of the ideal person according to Kant; if anything, the main feature is autonomy understood as self-governance guided by universal principles. Any attempt to divorce Kant's notion of autonomy from his view of morality, for example by understanding it solely in terms of 'neutral' self-control, is bound to be untrue to Kant's philosophy.

I will qualify this claim now on the basis of other of Kant's writings. Firstly, I believe something more needs to be said about what Kant means by self-control and the role it plays as part of his ideal of autonomy. Secondly, in the following section on Kant's answer to Problem Three, I will qualify my claim about the absence of any 'neutral' ideal of autonomy in Kant's work.

8.2.5. Is self-control a Kantian virtue? More on Kant's second-best ideal of the person

I have already alluded to the Kantian ideal of holiness. But, as noted, Kant also defends a less demanding ideal, and an ideal which needs to be included, I think, in any Kantian version of the extended ideal of the autonomous person. We have construed the ideal of autonomy as (among other things) inner order and harmony. This, it has been argued, is the proper understanding of self-mastery, and in any case of inner autonomy, and contrasts with self-control as the repression of desires. Kant's description of a virtuous character contains elements of both. On the one hand, there is the language of control and constraint. In the *Metaphysics* Kant argues that since men are 'rational *natural* beings, who are unholy enough that pleasure can induce them to break the moral law, even though they recognise its authority', when they obey the law they do it reluctantly. They are constrained. But since they are also *free* beings, 'the constraint that the concept of duty contains can be only *self-constraint*', and it is in this way that man's constraint (or *necessitation*) can be 'united with the freedom of

his capacity for choice' (Me 379–80). (Note that this is, in a nutshell, Kant's way of solving Problem Two: he explains precisely how and why freedom is compatible with dependency.) Such a self-constraint requires moral strength. Kant writes:

> *Virtue* is the strength of man's maxims in fulfilling his duty. Strength of any kind can be overcome, and in the case of virtue these obstacles are natural inclinations, which can come into conflict with man's moral resolution; and since it is man *himself* who puts these obstacles in the way of his maxims, virtue is not merely a self-constraint (for then one natural inclination could strive to overcome another), but also a self-constraint in accordance with the principle of inner freedom, and so through the mere representation of one's duty in accordance with its formal law. (Me 394)

This self-constraint is, according to Kant, commanded by reason, since '… reason says, through the concept of virtue, that one should *get hold of* oneself. […] Since virtue is based on inner freedom, it contains a positive command to a man, namely to bring all his capacities and inclinations under his (reason's) control and so to rule over himself' (Me 408). This, then, is how Kant understands self-rule, autonomy.

We may conclude that Kant's ideal of the autonomous person is one who struggles against his inclinations, a struggle which is at the same time a struggle for freedom as Kant understands it. However, this is not the complete picture, and not correct as a comprehensive representation of Kant's doctrine of virtue and his ideal of the autonomous person. In fact, the real ideal of the person Kant defends is a condition in which there is no longer a real necessity for self-control or resistance of temptation. Kant suggests this at some point: 'Virtue so shines as an ideal that it seems, by human standards, to eclipse *holiness* itself, which is never tempted to break the law' (Me 397). A genuinely virtuous character, then, is 'someone who feels little or no temptation in the first place rather than someone who is engaged in a constant and heroic struggle with temptation. It is in turn this lack of openness to temptation to do otherwise that makes it possible to be cheerful in the performance of duty' (Allison 1990: 163). This interpretation is consistent with Kant's view of the compatibility of freedom and constraint. In a footnote Kant claims that 'the less a man can be constrained by natural means and the more he can be constrained morally […], so much the more free he is' (Me 382). According to Kant, we are free 'in the highest degree' if we are 'unable to resist the call of duty' (Me 382). This inability can be understood as a lack of openness

to temptation; a true volitional necessity that does not have its source in 'what we care for', as Frankfurt would have it, but in duty.

In short, there are at least three ideals here: (1) holiness, (2) heroic struggle with temptation, and (3) not being open to temptation, striving for holiness, being cheerful in the performance of duty. The third one gets very close to the ideal of holiness (as I quoted earlier: 'Virtue so shines as an ideal that it seems, by human standards, to eclipse *holiness* itself, which is never tempted to break the law' (Me 397)), but is still distinct since it is meant as an ideal for humans, in other words, since holiness is striven for but never reached. This means, in my view, that there is a comparative aspect to the ideal: who is nearer and nearest to the ideal of holiness? I take Allison to allude to this when he writes that 'the truly virtuous are those who do not allow themselves to be tempted or, more properly, since no finite agent is beyond the possibility of temptation, those who do not allow themselves to be tempted by the things that are irresistible to the rest of us' (Allison 1990: 164). However, Kant argues for the virtue of *humility*. Surely, 'man's greatest perfection is to do his duty *from* duty' (Me 392) but he warns that 'trying to equal or surpass others in this respect, believing that in this way one will get an even greater inner worth, is [a kind of] *ambition* (*ambitio*), which is directly contrary to one's duty to others' (Me 435). True humility, he argues, follows not from comparison of ourselves with others but rather from 'our sincere and exact comparison of ourselves with the moral law' (Me 436).

The notion of comparison raises the question of measurement. We noted earlier Kant's view that 'strength of any kind can be recognised only by the obstacles it can overcome' and that, according to Kant, 'it is man *himself* who puts these obstacles in the way' (Me 394). Does this mean, then, that we should (not) create our own inclinations? Allison writes:

> The point, of course, is not that one creates one's inclinations but rather that one allows them to become obstacles to morality by placing a higher value on their satisfaction than is placed on the fulfilment of duty. Self-control, then, must be understood as control over this propensity rather than merely over the inclinations themselves. (Allison 1990: 164).

Self-control understood in this way is not the repression of desires, but rather something on a different level, namely the proper level on which the ideal of autonomy operates. The ideal of autonomy is not

directly concerned with the control of this or that desire, but rather with the independent and self-directing process of moral judgement which decides between desires, reasons, and values. In Kantian terms, this means the judgement concerning whether to follow one's duty or not, which generates the problem of which criteria we use to decide between duty and inclination. This is the kind of problem a theory of the ideal of the autonomous person needs to deal with, and this makes it into proper moral psychology (necessary for the discussion of the ideal of the autonomous person) as opposed to theories about the repression of desires.

Kant makes a distinction between actual strength of character and mere capacity for self-control. The latter is common to all rational agents; the former 'must be acquired through a process of self-discipline' (Allison 1990: 164). This distinction corresponds to the distinction between the capacity and the condition of autonomy as relevant to the ideal of the autonomous person (see Section 1.5.3.). In particular, 'the rules for practising virtue (*exercitiorum virtutis*) aim at a frame of mind that is both *valiant* and *cheerful* in fulfilling its duties' (Me 484). Kant admits that this involves 'sacrificing many of the joys of life, the loss of which can sometimes make one's mind gloomy and sullen' (Me 484). However, he also says that the 'self-torture and mortification of the flesh' some ascetics engaged in are not directed to virtue at all (Me 485). What does Kant mean by self-discipline then?

By a process of self-discipline Kant doesn't understand a sort of conditioning, a 'merely habitual training in "goodness",' but a training which develops 'the capacity for independent moral judgement on the basis of firmly held principles' (Allison 1990: 165): 'What distinguishes virtue from other forms of self-control for Kant is that it is based on a "principle of inner freedom", that is to say, a moral principle freely adopted by the agent' (Allison 1990: 164–5). Inner freedom refers to 'the capacity for self-constraint [...] by pure practical reason' (Me 396). Furthermore, I have already referred to Kant's argument that only in this way can necessitation be united with freedom. That a moral principle be 'freely adopted' is essential to making this virtue part of the ideal of the autonomous person. Furthermore, this autonomy concerns moral judgement as opposed to other forms of self-control. The ideal of the person that emerges is that of an agent freely adopting principles as an exercise in moral judgement. For Kant, this means ultimately adopting the principle that one should act from duty alone: 'Man's greatest moral perfection is to do his duty *from duty* (for the law to be not only the rule but also the incentive of his actions)' (Me 392).

If we lack this virtue, Kant does not think we are vicious or wicked. That would be a deliberate and 'principled violation of one's duties', which 'does not mean that not doing one's duty is itself made into a principle (that would be the mark of a diabolical will), but rather that the vicious person is firmly committed to immoral principles' (Allison 1990: 168). It may be that we are not vicious in this sense and that we still lack virtue. As Kant writes in *Religion*, it may be that our will suffers a 'lack of sufficient strength to follow out the principles it has chosen for itself' (Re 32).

But whatever the refinements that can be made within not striving for the ideal of the autonomous person, the ideal itself is clear. The vicious person who is firmly committed to immoral principles and the person with a diabolical will are following principles and are self-directing. And the one who has chosen a principle but does not follow it has been self-directing too. However, we do not want to make self-direction a sufficient condition for the ideal of the autonomous person. The ideal needs to include reference to being moral; otherwise it is perhaps a *sort* of 'autonomy', 'independence', 'freedom', and 'self-direction', but not of the sort that we want to hold up as an ideal to aspire to.

8.3. Kant's answer to Problem Three

8.3.1. Two contradictory positions on the relation between autonomy and morality

I would like to qualify the conclusion arrived at in the previous sections on the relation between (Kant's view of) morality and Kant's ideal of autonomy, in particular the conclusion of the section on 'Autonomy'. Although in the *Groundwork* Kant equates autonomy with obedience to the law, and therefore strongly connects autonomy with morality, this view is not sustained throughout his writings. It has been argued[54] that in his later writings Kant abandoned this view of autonomy in favour of a (more) 'neutral' notion of autonomy with the introduction of the *Wille/Willkür* distinction. I will say more about this distinction soon (in this section and Section 8.3.2.) but, in brief, *Wille/* is the part of us that participates in making the norm; the function of *Willkür* is to choose in the light of this norm. *Willkür* is the will considered 'merely in terms of its radical capacity of free choice,' whereas *Wille* 'refers to the purely rational aspect of the will' (Silber 1960: ciii–civ). On the basis of this distinction, it is argued, Kant can explain 'how freedom to do evil, and, indeed, choice between morally indifferent alternatives are possible' (Allison 1990: 95). This is Kant's answer to

Problem Three of the extended ideal developed in the first part of this book: by making the distinction between *Wille* and *Willkür* Kant allows for the intuition that we may be autonomous persons, in Kant's terms we may have an autonomous will (corresponding with *Wille*), but still fail to choose the good (corresponding with a failure of *Willkür* to choose in the light of the norm *Wille* made). In other words, the Kantian concept of *Willkür* leaves freedom to choose evil. Choosing evil, then, can be seen as a failure of *Willkür* to 'do its job' or a failure of *the person* to let *Willkür* perform its proper function. I'm not sure which of these two interpretations is properly Kantian. But whatever interpretation is right, it is clear that with these concepts Kant seems to be able to account for the fact that a person may possess, in principle, an autonomous will, and still do evil.

However, if this ideal of autonomy (among other things) includes that we always have the choice between good and evil, on what grounds do we make that choice? Kant, in contrast to most of the other philosophers discussed in this book, has to be given credit for (1) considering the question and (2) attempting to provide a satisfactory answer to it. I will now show that The *Wille/Willkür* distinction and his doctrine of radical evil can be interpreted as such. I start with examining what follows from the *Wille/Willkür* distinction and the view of autonomy that emerges from it.

If Kant claims in his later writings that we are still capable of choosing evil, then the following questions emerge. Firstly, on what basis do we make the choice between good and evil? Secondly, is this freedom to choose evil an instance of autonomy? According to Allison, in this view 'even heteronomous willing turns out to involve a certain kind of autonomy' (Allison 1990: 95). In short, it now appears that Kant defends here a morally neutral conception of autonomy. Thirdly, is this morally neutral conception of autonomy not in contradiction with his earlier conception of autonomy in the *Groundwork*? According to Allison, 'the major problem with this second view is that it undermines the distinction between autonomy and heteronomy, since all rational agency is "autonomous" in *this* sense' (Allison 1990: 96).

In other words, if the *Wille/Willkür* distinction is taken to make a difference between *Wille* as the maker and follower of the law, on the one hand, and *Willkür* as a term to refer to possibility of choice for evil rather than good, Kant's view of autonomy (as including the freedom to choose evil) becomes deeply problematic. Then there are two possibilities: either we interpret the *Wille/Willkür* distinction differently so as to avoid a contradiction with the doctrine of autonomy in the *Groundwork*

(let's call this the *first option*); or we dismiss the *Wille/Willkür* distinction altogether *because* of the contradiction with what I take Kant's core idea of autonomy (let's call this the *second option*) to be. The second option could be preferred on the basis of the following reasons. Firstly, could the *Wille/Willkür* distinction be possibly interpreted differently? (This is meant at this stage as a rhetorical question; the answer I suggest is 'No'. However, I will reconsider another answer to this question later.) Secondly, is it not the only option fit to uphold the ideal of the autonomous person as defended by Kant in the *Groundwork*? Consider the following argument for this claim.

Kant's purpose is not to defend autonomy as another term for freedom in the sense of being able to choose between good and evil, or being able to choose at all. Autonomy, in Kant's view, is rather defined by *determination*, in the sense that I determine the law and the law determines me. It is independence, perhaps, but not independence to choose between good and evil. Rather, it is self-rule in the sense of giving the law to oneself: I let myself be determined by the law. And since the moral law is by definition good, I let myself be determined by the good, I give the good to myself. It is also not independence from 'causal determination by one's needs as a sensuous being', nor is it 'total freedom from these needs', since we are not divine or holy but human (Allison 1990: 97). We do have needs, and they influence us one way or other. According to Allison, Kant means *motivational independence*: 'a capacity for self-determination independently of, and even contrary to, these needs' (Allison 1990: 97). In other words, we can have reasons to act that are independent of our needs as sensuous beings. Kant understands these reasons primarily as principles: 'to attribute the property of autonomy to the will is to attribute to it the capacity to be moved to action by a rule of action (practical principle) that makes no reference to an agent's needs or interests as a sensuous being' (Allison 1990: 98). Such a notion of autonomy is not morally neutral, at least in the following senses. Firstly, it clearly favours acting for the sake of (moral) principles. It is not true that if and when we are autonomous (or if our will is autonomous) we have, according to Kant's notion of autonomy in the *Groundwork*, still any choice between duty and inclination, or between good and bad. Kant writes that 'the principle of autonomy is "Never to choose except in such a way that in the same volition the maxims of your choice are also present as universal law"' and that this practical rule is 'an imperative – that is, that the will of every rational being is necessarily bound to the rule as a condition' (Gr 440). On this view, it is also not possible to be autonomous

only to a certain extent. Autonomy and morality are intimately connected in Kant's theory: (1) Kant argues that autonomy is a necessary condition for the possibility of morality, and (2) Kant elevates autonomy of the will as the supreme principle of morality itself.

Firstly, Kant sees autonomy as the necessary condition for the possibility of morality: 'the thesis is that a will and only a will with the property of autonomy is capable of acting on the basis of a categorical imperative' (Allison 1990: 99). Allison rightly notes that Kant's dichotomy between autonomy and heteronomy is a mutually exclusive model of volition: 'either the will gives the law to itself, in which case we have autonomy, or the law is somehow given to the will from without, in which case we have heteronomy' (Allison 1990: 99).

Secondly, Kant argues – as considered earlier – that autonomy is also one of the formulas of the categorical imperative itself, and therefore an ethical principle itself. Kant argues that '*morality* lies in the relation of actions to the autonomy of the will – that is, to a possible making of universal laws by means of its maxims. An action which is compatible with the autonomy of the will is *permitted*; one which does not harmonise with it is *forbidden*' (Gr 439). When explaining autonomy of the will Kant uses the title 'Autonomy of the Will as the Supreme Principle of Morality', claiming that 'the principle of morality must be the categorical imperative, and that this in turn commands nothing more nor less than precisely this autonomy' (Gr 440). This means that autonomy is given a much more important role than just a necessary condition for the possibility of morality.

Note that this 'solution' (the second option I try to argue for now) does not exclude all choice. It only says that *if* I am autonomous, I am moral/good. Although we all have the capacity of autonomy (according to Kant), we may not exercise this capacity. However, the problem with this argument is that it does not solve Problem Three, since it is not clear on what grounds I may decide (not) to exercise my capacity of autonomy. In other words, even if we take the second option, setting aside the *Wille/Willkür* distinction, we are *still* confronted with the problem of the possibility of choice (for evil) and the grounds for this choice. Kant attempts to deal with this problem with his account of radical evil, which we will discuss in the next section.

8.3.2. The *Wille/Willkür* distinction reconsidered: Kant's concept of radical evil

In the above discussion I have argued for the *second option*, defending the view that the *Wille/Willkür* distinction, and the view of

autonomy that emerges from it, contradict the ideal of autonomy in the *Groundwork*, that they cannot be reconciled, and that therefore we would do better to disregard the *Wille/Willkür* distinction as a part of the Kantian ideal of autonomy. But perhaps I have dismissed the *first option* too soon? Is a reconciliation not possible? Could the *Wille/Willkür* distinction not indeed be understood differently? I will now re-open that discussion.

According to Allison, Kant's motivation for introducing this distinction is 'to clarify his conception of the will as *self-determining* and, ultimately, as autonomous' and 'presupposes a certain duality of function within the will' (Allison 1990: 130). The function of *Wille* is to provide the norm; the function of *Willkür* is to choose in the light of this norm. The question is then: What does *Willkür*'s 'choice' mean? If the norm is provided, is there still any choice to be made? If Kant affirms that *Willkür* is 'free', what does this freedom mean? If Kant writes that one should not try to 'define freedom of *Willkür* as the power to choose between the alternatives of acting with or against the law' (Kant in the *Metaphysics of Morals* as quoted in Allison 1990: 133), how *should* one then understand the freedom of *Willkür*? It is clear from my discussion of the *Wille/Willkür* distinction that Kant (too) struggles with this view that we can still choose between following the law or not, and ultimately between good and evil. It is a view easy to arrive at: the modern ideal of the autonomous person includes the freedom to choose, and it is only one step further to take this choice to include the choice between good and evil. If I am autonomous in the sense of having the freedom to direct my own life, why can't I choose for myself between good and evil? Should I not incorporate this view in the extended ideal I am trying to develop? On the grounds that there are many problems with the 'freedom to choose between good and evil' view, I could remain with the Platonic interpretation and say that if we *are* autonomous, we are simply 'determined' by the good, and there is no question of choice. However, as we see in Kant, such a view seems to be unacceptable to the modern mind. We want to have the freedom to choose between good and evil – even if such a notion of freedom is not coherent and ought (therefore) not to be part of a notion of autonomy. But if it is not a coherent notion of freedom, can it be a freedom we really want? And if I know the good, why would I want to choose evil? So why would I want the freedom to choose evil? In the rest of this book, I can only show that one can go two ways, and that the two ways are equally problematic, but for different reasons. The 'determinism' view of autonomy is problematic since it contradicts

our modern desire for (this specific) freedom at all cost – and therefore the desire for this kind of freedom to be part of our ideal of autonomy. The 'freedom' view of autonomy is problematic since it is difficult to see on what basis one can make the choice and why one would want to choose evil at all. (Note that 'determinism' and 'freedom' are meant here as very specific terms referring to aspects or possible versions of an ideal of *autonomy* and have therefore a different meaning than their use in relation to the common freedom/determinism distinction.)

Kant's discussion of radical evil can be seen from the perspective of this discussion. In an attempt to hold on to the 'freedom' view, Kant feels compelled to explain why we would still want to choose evil by saying in the *Religion* that there is 'a *radical* innate *evil* in human nature' (Re 28), more, that 'man is evil by nature' (Re 27). Needless to say, this view creates a lot of problems. But let us first look at what Kant means by 'radical evil'. Kant doesn't mean a certain form of evil, but rather 'the root or ground of the very possibility of all moral evil' (Allison 1990: 147). According to Kant, evil itself consists in the adoption of maxims contrary to the law: 'Man is *evil*, can mean only, he is conscious of the moral law but has nevertheless adopted into his maxim the (occasional) deviation therefrom' (Re 27). But, if our original predisposition 'is a predisposition to good', as Kant claims (Re 38), why would we choose evil? Is there any ground for choosing evil? This is precisely what we discussed earlier: Kant's concept of radical evil can be seen as an affirmation of the claim that there *is* a reason why we would want to choose evil. Radical evil is the ground of 'the possibility of the adoption of immoral maxims' (Allison 1990: 147). Kant's argument is that we have to assume such a ground to account for the possibility of any moral evil. Kant seems to want to account for the observation that in reality humans do *not* have a good or holy will, that we are also 'creatures of desire and inclination, which, as resting on natural causes, are neither completely in our control nor necessarily in agreement with the dictates of morality' (Allison 1990: 156). As noted already, Kant holds the view (in the *Metaphysics*) that men are not only rational but also *natural* beings, 'unholy enough that pleasure can induce them to break the moral law, even though they recognise its authority' (Me 379); in other words, they can be induced to evil (according to the definition of evil given above). However, does this observation entitle Kant to conclude that there is such a thing as *radical* evil? That there really is a ground for doing evil? We may be not capable of holiness, but surely from this observation it does not follow that we have 'an actual propensity to subordinate moral considerations

to our needs as sensuous beings, that is, a tendency to let ourselves be tempted or "induced" by inclination to violate the moral law even while recognising its authority' (Allison 1990: 157). And even if we did have such a 'propensity', it is not clear to me why this necessarily should result in *evil*. Firstly, it may be that we resist the temptation, and secondly, there are degrees of violating a moral law.

8.3.3. Conclusion

Although Kant presents us with a serious attempt to deal with Problem Three, my discussion of this attempt shows that *any* answer to this problem remains problematic. To say that morality and autonomy are so intimately connected that there is no ground for the choice of evil seems to contradict the intuition we may have that I can be autonomous and still choose evil. However, I have also shown that to support this intuition with arguments runs inevitably into the question of what ground there is for choosing evil. Kant tries to answer this question but his answer appears unsatisfactory. In conclusion, I doubt whether any satisfactory answer can be given – except perhaps the answer that there is no ground, no radical evil. Which brings me back to the start of the argument, which is to say that morality and autonomy are inextricably connected.

Whatever the problems with Kant's discussion of radical evil are, I think it is clear which ideal of the person emerges from it: the person who is not tempted by his needs as a sensuous being to violate the moral law. But, as the discussion of various Kantian ideals has shown, this does not mean that Kant expects us, as human beings, to be able to reach this state of non-temptation, this state of *holiness*. As I have shown, the ideal person according to Kant may be either the one who has a perfectly good or holy will, or – if this is not possible – the person who does his duty *gladly* – not reluctantly. But the doctrine of radical evil accounts for the fact that even this 'second-best' ideal is difficult to reach. We often *are* reluctant. Allison writes that 'the doctrine of radical evil not only defines our moral condition but also sets the moral agenda for finite, imperfect beings such as ourselves, namely, to struggle to the best of our ability against an ineliminable reluctance to subordinate the requirements of our sensuous nature to the dictates of morality' (Allison 1990: 162). In so far as this moral agenda can be construed as an ideal of the person, this means that we, as persons, *ought* to struggle against this reluctance. And although this may be called Kant's 'third-best' ideal of the person, it is still a very demanding one – certainly more demanding than the requirement not to do evil.

8.4. Conclusion

8.4.1. Kantian autonomy and the extended ideal of autonomy

In conclusion, to the extent that Kant argues that autonomy of the will 'is unavoidably bound up with [morality] or rather is its very basis' (Gr 445), his argument supports my efforts to consolidate the extended ideal of the autonomous person, an ideal which includes a strong connection between autonomy and morality as opposed to interpretations of autonomy as a morally 'neutral' concept. This does not mean that we *need* to take the (whole) Kantian framework on board; rather, the Kantian notion of autonomy provides a possible answer to questions that flagged up earlier in this book regarding the relation of the autonomy to reason and morality.

However, just as was the case with Plato and Augustine, a discussion of autonomy using Kantian materials involves a confrontation with a whole metaphysics which we do not necessarily want to adopt. We notice that Kant's discussion is based on a distinction between (man as member of) an intelligible world and a sensible world:

> The moral 'I ought' is thus an 'I will' for man as a member of the intelligible world; and it is conceived by him as an 'I ought' only in so far as he considers himself at the same time to be a member of the sensible world. (Gr 455)

Furthermore, my discussion of the Kantian ideal of autonomy in relation to the extended ideal of autonomy – in terms of autonomy by self-guidance – involved costly metaphysical commitments. Since a belief in such commitments cannot be taken for granted, we may question whether there is such a thing as 'good' or 'principles' and whether it makes sense to speculate about their relationship.

Another example of a costly metaphysics is that Kant's contrast between duty and inclination suggests a dualistic view of human nature: 'inclination must be construed in a broad sense to refer to any stimulus to action that stems from our sensuous, as opposed to our rational, nature' (Allison 1990: 108). This dualistic view of human nature (the oppositions duty/inclination and rational/sensuous) corresponds with a dualistic view of the world (intelligible/sensible). To repeat the quotation I gave above: 'The moral "I ought" is thus an "I will" for man as a member of the intelligible world; and it is conceived by him as an "I ought" only in so far as he considers himself at the same time to be a member of the sensible world' (Gr 455).

The dualistic view of human nature is shared by the Platonic and Christian self-understanding. Frankfurt's hierarchical model of volition, on the contrary, suggests a volitional monism, attempting to avoid the question whether 'inclination' or 'duty', the 'sensuous' or the 'rational' should prevail. By saying that a higher-order desire takes priority over a lower-order desire it avoids saying something about whether it is *good* to have this or other desire. The extended model in this book, on the contrary, does say that morality *is* essential to my *autonomy*.

8.4.2. General conclusion

My conclusion is that if we want to have a coherent ideal of the autonomous person, we will not be able to avoid making some costly metaphysical claims. We might want to rework Plato, Augustine, and Kant in terms that are more acceptable to our contemporary eye, but broad notions such as 'the good', 'the person', 'morality', etc. cannot be avoided if we want to make sense of autonomy. It is not a coincidence, I think, that currents of thought wanting to abolish the notion of the subject, for example, destroy at the same time the notion of autonomy and indeed the notion of morality itself. (I am thinking here about the so-called 'post-modern' currents of thought, but I do not have room here to elaborate this point.)

A major problem that figured in this discussion was that of how to decide between – in Kant's terms – doing your duty and following your inclinations. It is not enough to observe that there is a 'natural dialectic' between the two, and to know that the problem arises from me being a part of both the intelligible and the sensible world. If I, as an individual, am faced with a specific conflict between duty and inclination, I would like to know how to resolve this conflict. When do I have to consider myself a member of the intelligible world, when of the sensible world? This is a problem for moral philosophy, but it does not need to be a problem for thinking about the ideal of the person. It has become clear from Kant's account of autonomy that to do your duty, to consider yourself a member of the intelligible world, etc. *is*, according to Kant, what it is to be autonomous. Therefore, we can be satisfied with saying that this is the ideal. I am not concerned with the moral question whether we ought to aspire and live up to the ideal of the autonomous person. Rather, I'm saying that if you do aspire to be autonomous, you will face certain problems. And these include, of course, the problem that it is difficult to make coherent sense of what the intelligible world is. Kant's idea refers to 'a "something" that

remains when I have excluded from the grounds determining my will everything that belongs to the world of sense,' to a 'more' beyond the world of sense; 'yet with this "more" I have no further acquaintance' (Gr 462). If we want to hold on to Kant's strong connection between autonomy and morality, our inquiry into the content of the ideal of the autonomous person faces the same limit as that of moral inquiry itself: what can we know about the good (Platonic version), God (Augustinian version), or the intelligible world (Kantian version)? The only thing we can say is that we need to say something about these 'metaphysical' implications since in the absence of any element of this kind the ideal of autonomy is empty and does not make sense. Nothing said here forces anybody to accept the existence of 'the good' (or 'God', or an 'intelligible world'), or to strive for the ideal of autonomy. But my argument is that you can't have just one of them.

Conclusion of Part II

To explain what Kant calls 'radical evil' we needed to consider complex notions involving 'good' and 'evil'. It became clear throughout the discussion that Kant could only be successful because he relied on an extravagant metaphysics at least as elaborate as that of Plato or Augustine. Kant's solution to the problems, therefore, left Problem One unresolved; he certainly did not achieve metaphysical economy.

We may prefer not to take on board the 'encumbering' metaphysical luggage that comes with Kant's ideal of autonomy. But is there, if we want to preserve the ideal of autonomy, a way between a costly metaphysics and the existentialist anti-metaphysical position? Is there a way between metaphysics and despair? Frankfurt's and Hill's view of autonomy deserve credit for attempting precisely such a 'third way'. But, as I have shown, they face problems and limitations that cannot be overcome within their own frameworks as they stand.

My conclusion is that searching for a coherent ideal of autonomy between a metaphysically 'compromised' ideal of autonomy including notions such as 'God' or 'good', on the one hand, and an existentialist ideal of autonomy drawing the consequences from not assuming such a metaphysics, on the other hand, is bound to remain unsatisfactory. To the extent that we accept the dominant modern ideal of autonomy, it seems that we must also accept some metaphysical views we might have preferred to avoid. Until an adequate morally 'neutral' and metaphysically 'economical' ideal of autonomy can be developed, we have to accept that the modern ideal of the autonomous person necessarily involves a strong moral component and a rich metaphysical framework. At least, that is so if we want to aspire to a coherent ideal of the person, one that makes sense, and we have many reasons to want to do so.

This does not mean, of course, that we have to accept those views uncritically. Furthermore, if we do not want to adopt the 'old' metaphysics of Plato, Augustine, or Kant, we may feel challenged to search for a 'new' metaphysics appropriate to the defence of the cogency of the ideal of autonomy. If this option is preferred, my book should provide some guidelines to constrain such a search, at least if it is to say something about autonomy. There may be other ideals worth considering, but if autonomy is to remain a dominant ideal in our society, we might want to question whether and to what extent this ideal is consistent with our metaphysics and the beliefs we hold about (the role of) metaphysics (in modern philosophy and society).

Notes

1 In particular in the chapter titled 'Autonomy'. The chapter is reprinted in Christman, J. (ed.), *The Inner Citadel* (1989).
2 Autonomy can refer to the capacity, the condition, the ideal, or the sovereign authority of self-government and independence (see Feinberg 1986: 28). I will say more about the distinction between capacity and condition later (see Section 1.5.3.).
3 I choose not to use the female pronoun for ease of reading. I noticed that the use of 'he or she', 'his or her' etc. makes the text less readable. However, it is understood that here and elsewhere in the book I mean to refer to both men and women when writing about persons or human beings.
4 Hereafter I shall refer to 'second-order desires' rather than 'second-order volitions', since for my purposes in this section this distinction, by itself, is not important; both terms imply the same ideal of the person. Moreover, in terms of Frankfurt's own account, second-order volitions can be interpreted as a higher-order desire (the effective desire to make a certain desire my will). I will argue this in Section 3.1.
5 To possess autonomy as a capacity is, of course, a necessary condition to achieve the condition of autonomy. But it also seems reasonable to say that the capacity is not a sufficient condition to achieve an actual state or condition of autonomy. With 'capacity' Feinberg refers to 'the ability to make rational choices', a 'competence' necessary for self-governance. But to be able to exercise this capacity is another matter. According to Feinberg, we do not only want autonomy in itself, but also 'its fruits – responsibility, self-esteem, and personal dignity' – and therefore when we aspire to autonomy as an ideal we also want to have the opportunity to actually govern ourselves. I do not have this opportunity 'if you overpower me by brute force and wrongfully impose your will on mine'; if, in general, circumstances beyond our control prevent us from enjoying *de facto* autonomy. (I will discuss this problem further later as a problem of (lack of) 'outer' autonomy.)
6 For example, the personal autonomy of the patient is interpreted as 'doing what you want with your body': absolute freedom of action (upon your body, by yourself or by others with your permission) within the limits of biological/physical laws and medical possibilities (not fixed but if possible adapting to your wishes). Similarly, the personal autonomy of the consumer is interpreted as 'doing what you want with your money': absolute freedom of action (buying) within the limits of your budget (although this is not fixed) and supply (certainly not fixed but rather adapting to your wishes). In both examples, the problem I want to point to in the context of my argument (there are many more) is that the person's *wishes* or *desires* are taken as given and as not-to-be-restrained, without asking further questions such as 'What does it mean to restrain myself?', 'What is the "I" in the sentence "I wish" or "I exercise restraint",' 'Which desires do I want to have?,'

'Is what I desire also what should be?' etc. In other words, the question about identity and the normative question are avoided. In the course of my book I will point to the need to raise these questions if we want a cogent ideal of autonomy. I shall show that if we say 'I wish to rule myself, to be the master of my life', etc. we need to think about what it is to do that and how we can do that, and that absolute freedom makes autonomy impossible rather than being its definition.

7 See, for instance, Mill, *On Liberty*, Ch. 3 and Ch. 5; Hume, *Treatise of Human Nature*, Book 3; and Hobbes, *Leviathan*, Ch. 21.

8 This discussion runs through the whole book; see for example Sections 3.6., 6.2.7., and 6.3. The matter is raised at several other points too.

9 Note that the distinction between concept and conception is not new and is not only applicable to the concept of autonomy. For example, in *A Theory of Justice* Rawls employs a distinction between the concept of justice and various conceptions of justice (the distinction is set up on page 5).

10 In quotations, the term 'the Good' may appear. Here and elsewhere I will use lower-case letters for 'the good', but the typographical difference is not meant to amount to a difference in sense.

11 Note that I use here 'the beautiful', 'the good', etc. as if we can simply replace the one by the other. This is not so; at least I do not want to suggest that they are the same. However, as I interpret Plato, they are replaceable with each other in regard to their function as the source of inspiration and madness. Here and in the rest of my book I will mainly focus on 'the good', since it plays a key role in my argument for compatibility in Chapter 4. I'm not interested – here and in Chapter 4 – in 'the good' as such, but in its role in Plato's ideal of the person and especially in its relation to *autonomy*. This relation will become clearer in Chapter 4. I will leave aside questions concerning the relation between autonomy and 'the beautiful', or between autonomy and 'the truth'. I will argue in Chapter 4 that although Plato writes about the madness related to the beautiful, his account can, by analogy, be transferred to the good.

12 Obviously there still are, and have been, (many) instances of 'scapegoat rituals', in traditionally 'Christian' societies or groups too.

13 Note that both Plato and Augustine (as we will see) view (doing) evil as (action due to the) absence of the good rather than ascribe to evil a positive reality.

14 The first number refers to the Book, the second to the page of the translation I used.

15 Dilman appears to have in mind the following passage (quoted in Section 2.3.2.): 'When I chose to do something or not to do it, I was quite certain that it was my own self, and not some other person, who made this act of will [...]' (VII.3/136). However, this passage is from the *Confessions*, to which Dilman does not refer. I shall take it that this was his intended source. Unfortunately, since he fails to cite a source for the quotation he offers it is impossible to be certain.

16 Certainly Frankfurt writes that 'it is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will and of forming volitions of the second order' (Frankfurt 1982 (1971): 87; see also Section 1.3.), but nevertheless the focus of his whole account is on the volitional rather than the rational capacities of the person (Frankfurt 1988: viii).

17 When I use the term 'extension' I mean supplementation.
18 Note that if 'decisive identification' boils down to the formation of higher-order desires, Frankfurt's distinction between second-order desires and second-order volitions is untenable. The latter concept refers to wanting a desire to be your will and succeeding in having that will (see Section 1.3.). To succeed, Frankfurt could argue, the identification with the desire has to be 'decisive'. But if 'decisive identification' is merely the formation of a higher-order desire (for example, a second-order desire), such a second-order volition can only be interpreted as a second-order desire.
19 Watson argues that 'human freedom cannot be understood independently of the notion of practical reason or judgement, and that this notion is bound up with a distinction between desiring and *valuing*' (Watson 1982: 8). I agree with this argument and I will discuss this further in Part Two in relation to Frankfurt's later work (Chapter 6, in particular my objections in Section 6.4.1.). At present, however, it is my aim merely to draw the reader's attention to this major problem in Frankfurt's account. We will consider afterwards whether and if so how Frankfurt's model could be extended to account for this problem.
20 I will search for other candidates in the following sections of this chapter, in Chapter 4, and, in a different vein, in Part II (this search runs through the whole book). For further discussion of Frankfurt in particular, see Chapter 6 (especially Sections 6.3. and 6.4.2.).
21 Murdoch refers in particular to Hampshire, S. (1959) *Thought and Action* London: Chatto & Windus, 1982.
22 Note that Wolf writes 'normatively' and not 'morally'. 'Normatively' does not automatically imply 'morally'. The normative includes the moral, but not all normative claims are moral. However, I choose to explicitly mention 'morality' here, since I want to suggest already here a link between the problem of the modern ideal and the Platonic idea of the good. This will become clearer in Chapter 4, and at the end of that chapter I will also return to the normativity/morality issue as part of the conclusion of my argument of Part I. At this point, however, I merely want to question Wolf's use of the word 'sanity', and strictly speaking I could have written 'normativity' as well.
23 Note that there remains a major problem with this answer. Even if 'the good' is not an 'object', if I am dependent on it for my autonomy, how autonomous am I really? I will return to this problem at the end of this chapter. For now, I concentrate on the constructive movement in my argument with the aim of arriving at a reconciliation, a synthesis of ancient and modern ideals of the person. In Part II, this synthesis will in its turn become a thesis that will be challenged.
24 However, next to reason Platonic love and madness may help as well, as I suggested in the former section; they are, in their way, (most) excellent too in regard to their role in the achievement of autonomy. I will discuss the role of Augustinian love further on in this chapter. Although I recognise that there may be tension between reason, on the one hand, and love and madness, on the other hand, I will not discuss this tension further here, and, apart from what I will say about the relation between reason and *emotion* below, discuss the role of reason and love/madness separately.

25 Note that in the last sentence Dilman should have used 'he' instead of 'it'.

26 Note that this problem should not be put in terms of normative versus moral authority, but rather in terms of a problem of obedience. Consider the following argument. First, the problem of whether we have the choice of choosing evil cannot be solved simply by deciding on the *scope* of the moral. Rather, for the third problem to be a problem it is necessary to assume that the moral has an overriding claim, and that, in this sense, the 'scope' of the moral is wider than the mere normative: there is also the non-moral normative, but this does not have 'priority', so to speak. If the moral was just one of the many features involved in a certain action or decision, I might as well choose evil. The issue of 'why choose evil' arises only if you believe that there is something like the 'morally good' and that this morally good should direct your actions. Only then the question arises: if you know the morally good, why choose evil? Only if I accept the good as having moral authority, does the question of obedience arise: Do I have the choice to disobey the moral authority of the ultimate point of reference? Second, therefore, it would be wrong to present the problem as a problem of whether the ultimate point of reference has merely normative authority or also moral authority. The person who struggles with the 'third problem' already accepts its moral authority. Only then does the question arise: Do I, as an autonomous being and if I want to be (and remain) autonomous, (still) have a choice? Does autonomy include the choice between good and evil?

Can the problem be solved by analogy with the problem of civil disobedience? No. Civil disobedience might be morally acceptable if you can refer to some point of reference with moral authority, such as human rights. And to argue for human rights, you can refer to a 'higher-order' point of reference – an ultimate one if necessary. But if you disobey the authority of the highest, ultimate point of reference, what do you refer to? What is your ground? But don't we have a free will, and therefore the freedom to disobey? The problem starts all over again.

27 Note that I might not only be cruel to others, I might be cruel to myself as well, especially if my ideal of inner autonomy is combined with the ideal of self-control (as it often is in modern times as it is in Plato's and Augustine's work). 'When the self in control is a ruthless autocrat (King Reason) imposing order with an iron hand, the inner conflict is squelched only at great cost to elements of the self, and the presentation of rigid narrowness to the outside world. Self-control can be totalitarian repression, and self-discipline can become self-tyranny' (Feinberg 1986: 46). I will return to this problem of repression when discussing objections to Frankfurt's later work in Part Two (Section 6.4.1., in particular my objection to Frankfurt's ideal of wholeheartedness).

28 Note that in Sartre's view choices and actions amount to the same, at least in the sense that, in Sartre's view, if I say 'I choose X' but I perform Y, I have in fact chosen Y rather than X. According to Sartre it does not make sense to say 'I choose to be a painter but (due to this circumstance, etc.) I cannot paint, I cannot *act* as a painter.' According to Sartre, either you paint or you don't paint; you are nothing (also not a painter) except what you make of yourself, in other words what you *do*.

29 I take Sartre to mean values that pre-exist us, humans (in the sense of existing independently from us), or values that pre-exist any particular choice-situation.
30 Phenomenology involves describing consciousness, that is, describing how things are like for us. For Sartre (too), the question is: what is it like to exist as a conscious agent? Sartre's view of consciousness is, as I noted already in my discussion of Murdoch's objections, that of an isolated will: isolated from the world (including from the 'lump of being' we are apart from our will). His view of the (absolutely) free chooser can be understood as derived from this view of what (or rather how) we are.
31 From the way Sartre argues in *Existentialism and Humanism* it is clear that there was already a good deal of criticism around in Sartre's time and he gives the impression of being very well informed about the views of his opponents. However, he does not give adequate references. 'Communists' or 'Christians' are often the most accurate terms he uses. For a recent (analytical) account of Sartre's philosophy (including objections) see Gregory McCulloch's book *Using Sartre* (1994), based on *Being and Nothingness* and other earlier work of Sartre. I will refer to it again below since it includes discussion relevant to the issue of autonomy.
32 On the contrary, it could be argued that in *precisely* these moments I am myself: what constrains my will is *me*. Consider the famous remark of Luther 'Here I stand; I can do no other', as often quoted in the literature on this subject (see Chapter 6 on Frankfurt). 'Here I stand' is a condition or cause of 'I can do no other,' in the sense that it is my identity (the 'I' from 'Here I stand'), the person that I am, which is a (sufficient) condition for or the direct cause of the constraints ('I can do no other') to my will. This is how I choose to interpret Frankfurt, to whose concept of volitional necessity I will turn after my discussion of Sartre.
33 Note that Sartre has a similar problem with his theory of emotions. In his *Sketch for a Theory of the Emotions* (1939) he argues that we are completely responsible for our emotions since they are conscious *acts*; they don't come from 'outside' the person (for a discussion of Sartre's theory of the emotions, see for example Solomon 1981). This theory contradicts the phenomenological fact that sometimes we feel overwhelmed by our emotions, we feel that something happens to us. In other words, we sometimes feel that we passively undergo our emotions and sometimes even consciously struggle against them. This is a problem for Sartre.
34 I will take up this issue again in my discussion of Frankfurt and explain it more fully. I will show that although Frankfurt avoids Sartre's incorrect phenomenological description of choice by developing the concept of 'volitional necessity' – allowing him to take into account the fact that we feel certain options to be real options whereas we exclude other 'options' – he fails to recognise and discuss the normative and moral dimension of the introduction of this concept.
35 The assumption made here that we (sometimes) have to (be able to) judge others is not an obvious one. I will return to this issue in my discussion of Frankfurt. For now, note that Sartre himself talks about judging others without discussing why this is necessary: he simply assumes it, and, as I argue in the next paragraph, he has the problem that he can't ground his

notions of commitment and judging others in his own framework. The idea of groundless choice does not allow any evaluation.

36 Note that Frankfurt uses 'autonomy' without discussing its meaning. He also seems to often use 'freedom' and 'autonomy' as having the same meaning.

37 Against Frankfurt it is possible to object that if this is so then there won't arise a genuine dilemma or even a choice problem in the first place. If I know what I care about, and if this constitutes a really volitional necessity, as Frankfurt argues, then I know what I want and I know what to do. But I will postpone consideration of objections for later (Section 6.4.1.).

38 Note that this could be seen as Frankfurt's answer to the question about the ground for doing evil. He suggests that there is a ground to 'subordinate moral considerations to others', namely personal integrity and identity. In his view, it seems, betraying moral principles can be justified if by doing this you don't betray what you care about most. But is this a ground? I will return to this issue in Section 6.2.6. and also and especially in Section 6.4.1. when objecting to Frankfurt's claim that love and care can have normative authority.

39 See also my discussion of Augustine in Part I (Sections 2.3. and 4.2.). Both Frankfurt and Augustine have something to say on the role of love in relation to autonomy – even if their accounts are very different – but when dealing with Augustine I also limited myself to questions of autonomy rather than freedom or liberation.

40 There are times, however, when Frankfurt stresses precisely this 'external' aspect. For example, he writes that love requires a person '*to submit* to something which is beyond his voluntary control and which may be indifferent to his desires' (Frankfurt 1988: 89). But if love is 'part of myself', how *can* it be indifferent to my desires? How can I be indifferent to my own desires? And how can I submit to myself? Therefore, I prefer to hold on to his view that we can be 'overwhelmed' by part of ourselves, stressing the 'internal' aspect of volitional necessity. Of course this formula may be no less paradoxical (How can I be overwhelmed by myself?) but at least it does not suggest indifference or something 'external' I submit to.

41 Note also Dworkin's arguments in his famous paper 'Is More Choice Better than Less?' (1988).

42 Note that I use 'normative' here rather than 'moral', normative being, in my view, a broader notion which encompasses all 'oughts', including the moral 'oughts'. With my example of the committed Nazi I have stressed the question about *moral* authority as a special case of the question about normative authority, since I believe that this makes it plain that there is a need for considering the 'ought'.

43 See for example Piaget, J. and Inhelder, B. (1969) *The Psychology of the Child*, New York: Basic Books. However, for the purposes of this brief discussion I will rely on a summary of his view as presented by Duska and Whelan (1977).

44 See for example Kohlberg, L. (1984) *The Psychology of Moral Development*, New York: Harper & Row; see also Power, F.C., Higgins, A. and Kohlberg, L. (1989), *Lawrence Kohlberg's Approach to Moral Education*, New York: Columbia University Press. For this brief discussion I will rely on a summary of his view as presented by Duska and Whelan (1977).

45 See his essay 'Kantian Constructivism in Moral Theory' in *Journal of Philosophy* Vol. 77, 515–79. The idea is that individuals in the original position show 'rational autonomy' in the sense of not being guided by prior conceptions of justice or by personal characteristics of persons. In this way, the agents achieve 'full autonomy' when they act in accordance with the self-imposed principles that would have been chosen under conditions of 'fair' and 'neutral' choice. My argument in this book can be seen as directed in part against this idea of *not* being guided and of *neutral* choice which is supposed to be 'fair'.
46 For Frankfurt's view of the relation between autonomy and identity see Sections 6.2.2., 6.2.5., 6.2.8., and 6.3. For my objections to this view see Sections 6.4.1. and 6.4.2.
47 Note that Paton translated the German term 'Metaphysik' as 'metaphysic'. 'Metaphysics' is the correct translation. However, I use Paton's title here and in my bibliography to provide a precise reference.
48 Hereafter I shall refer to the *Groundwork* as 'Gr', to the *Religion* as 'Re', and to the *Metaphysics* as 'Me', followed by the page number.
49 The translation used is H.J. Paton (1948). I will use references to the standard edition issued by the Royal Prussian Academy in Berlin as given by H.J. Paton in the margin of his translation.
50 I will re-examine this suggestion later (Section 8.2.3.).
51 Page references to the Prussian Academy edition on which Mary Gregor's translation (1991) is based.
52 'Inclined to be inclined'…. This phrase suggests the possibility of infinite series of inclinations as well as duties. I have the duty to do my duty, and perhaps also the duty to do my duty to do my duty, etc. (I have already alluded to this problem in the previous section).
53 Compare this with the Christian idea of autonomy: if I act morally I am in fact truly disclosing myself.
54 See, for example, John Silber in 'The Ethical Significance of Kant's Religion' (1960).

Bibliography

Allison, H.E. (1990) *Kant's Theory of Freedom*, Cambridge: Cambridge University Press

Augustine, *Confessions* (trans. R.S. Pine-Coffin), London: Penguin Books, 1961

Augustine, *On Free Choice of the Will* (trans. Anna S. Benjamin and L.H. Hackstaff), Indianapolis: Bobbs-Merrill Educational Publishing, 1964

Berlin, I. (1958) 'Two Concepts of Liberty', reprinted in *The Proper Study of Mankind: An Anthology of Essays*, London: Chatto & Windus, 1997

Christman, J. (ed.) (1989) *The Inner Citadel: Essays on Individual Autonomy*, New York/Oxford: Oxford University Press

Dilman, I. (1999) *Free Will: An Historical and Philosophical Introduction*, London: Routledge

Dodds, E.R. (1951) *The Greeks and the Irrational*, Berkeley & L.A.: University of California Press

Duska, R. and M. Whelan (1977) *Moral Development: A Guide to Piaget and Kohlberg*, Dublin: Macmillan

Dworkin, G. (1988) 'Is More Choice Better than Less?', in Dworkin, G. *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press.

Euripides, *Medea, The Phoenician Women, Bacchae* (trans. J.M. Walton), London: Methuen, 1998

Feinberg, J. (1973) 'The Idea of a Free Man', in Doyle, James F. (ed.), *Educational Judgements*, London: Routledge & Kegan Paul

Feinberg, J. (1986) *The Moral Limits of the Criminal Law* (Vol. III): *Harm to Self* New York/Oxford: Oxford University Press

Frankfurt, H.G. (1971) 'Freedom of the Will and the Concept of a Person', reprinted in Watson, G. (ed.), *Free Will*, Oxford: Oxford University Press, 1982

Frankfurt, H. (1988) *The Importance of What We Care About*, Cambridge: Cambridge University Press.

Frankfurt, H.G. (1999) *Necessity, Volition, and Love*, Cambridge: Cambridge University Press

Girard, R. (1999) *Je vois Satan tomber comme l'éclair*, Paris: Grasset

Hampshire, S. (1959) *Thought and Action*, London: Chatto & Windus

Hill, T.E. (1989) 'The Kantian Conception of Autonomy', in Christman, J. (ed.), *The Inner Citadel: Essays on Individual Autonomy*, New York/Oxford: Oxford University Press

Hill, T.E. (1991) *Autonomy and Self-Respect*, Cambridge: Cambridge University Press

Jowett, B. (1953) 'Introduction' in: Jowett, B. (ed., trans.), *The Dialogues of Plato* (Vol. III), Oxford: The Clarendon Press

Kant, I. (1785) *Groundwork of the Metaphysic of Morals* (trans. H.J. Paton), London: Hutchinson, 1948

Kant, I (1793) *Religion within the Limits of Reason Alone* (trans. T.M. Greene and H.H. Hudson), New York: Harper & Row, 1960

Kant, I. (1797) *The Metaphysics of Morals* (trans. M. Gregor), Cambridge: Cambridge University Press, 1991

Kitto, H.D.F. (1961) *Greek Tragedy: A Literary Study*, London: Methuen
McCulloch, G. (1994) *Using Sartre: An Analytical Introduction to Early Sartrean Themes*, London/New York: Routledge
Murdoch, I. (1956) 'Vision and Choice in Morality', in Murdoch, I., *Existentialists and Mystics*, London: Penguin Books, 1999
Murdoch, I. (1961) 'Against Dryness', in Murdoch, I., *Existentialists and Mystics* London: Penguin Books, 1999
Murdoch, I. (1970a) 'The Idea of Perfection', in Murdoch, I., *The Sovereignty of Good*, London: Routledge & Kegan Paul
Murdoch, I. (1970b) 'On "God" and "Good"', in Murdoch, I., *The Sovereignty of Good*, London: Routledge & Kegan Paul
Murdoch, I. (1970c) 'The Sovereignty of Good over other Concepts', in Murdoch, I., *Existentialists and Mystics*, London: Penguin Books, 1999
Nietzsche, F. (1872) *Die Geburt der Tragödie*, in Nietzsche, F., *Werke in zwei Bänden (Band I)*, München: Carl Hanser Verlag, 1967; trans. S. Whiteside as *The Birth of Tragedy*, London: Penguin Books, 1993
Plato, *Phaedrus*, in Jowett, B. (ed., trans.), *The Dialogues of Plato* (Vol. III), Oxford: The Clarendon Press, 1953
Plato, *Republic*, in Jowett, B. (ed., trans.), *The Dialogues of Plato* (Vol. II), Oxford: The Clarendon Press, 1953
Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press
Ross, W.D. (1930) *The Right and the Good*, Oxford: Clarendon Press
Sartre, J.-P. (1939) *Esquisse d'une théorie des emotions*, Paris: Hermann; trans. P. Mairet as *Sketch for a Theory of the Emotions*, London: Methuen, 1962
Sartre, J.-P. (1948) *Existentialism and Humanism* (trans. P. Mairet), London: Methuen, 1973
Silber, J.R. (1960) 'The Ethical Significance of Kant's Religion', in Kant, I. (1793), *Religion within the Limits of Reason Alone*, New York: Harper & Row, 1960
Solomon, R.C. (1981) 'Sartre on Emotions', in Schilpp, P.A. (ed.), *The Philosophy of Jean-Paul Sartre*, La Salle, Ill.: Open Court
Taylor, C. (1976) 'Responsibility for Self', reprinted in Watson, G. (ed.), *Free Will*, Oxford: Oxford University Press, 1982
Velleman, J.D. (2002) 'Identification and Identity', in Buss, S. and L. Overton (eds), *The Contours of Agency: Essays on Themes from Harry Frankfurt,* Cambridge, Mass.: MIT Press
Watson, G. (1982) 'Introduction', in Watson, G. (ed.), *Free Will,* Oxford: Oxford University Press
Wolf, S. (1988) 'Sanity and the Metaphysics of Responsibility', reprinted in Christman, J. (ed.), *The Inner Citadel: Essays on Individual Autonomy*, New York/Oxford: Oxford University Press, 1989

Index