In Defence of Evaluative Compatibilism and Critical Evaluative Free Will Optimism

A Consequentialist Assessment of the Free Will Debate

Stijn van Gorkum (636669)

Bachelor Thesis

Supervised by:
dr. A.J.P.W. Dooremalen

Abstract

In this thesis, I will consider various aspects of the contemporary free will debate from a consequentialist perspective. In chapter 1, I will endorse a particular theory of what matters in life and a certain view of what moral rightness consists in. In chapter 2, I will provide an outline of all the various freedoms that have been proposed as (parts of) accounts of free will and of all the various responsibilities that have been proposed as (parts of) accounts of moral responsibility. In chapter 3, then, I will proceed to argue in favor of my first main thesis – that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms are compatible with determinism, and that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities are compatible with determinism (evaluative compatibilism). Finally, in chapter 4, I will argue in favor of my second main thesis – that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms exist in the actual world and that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities exist in our world, but that everyday and scientific observations do indicate that we have some valuable freedoms and responsibilities to a lesser degree than would be ideal (critical evaluative optimism). I do not expect to have convinced you, by the end of this thesis, that my ethical theories are the most intuitive ones; and neither do I expect to convince all of you that both (or either) of my main theses are (is) correct. What I do hope to convince you of, however, is that ethics is more important to the free will debate than many seem to think. For some ethical theories, the problem of free will simply does not have an edge. That is something we should all acknowledge.

Table of Contents

Abstract 2

Table of Contents 3

Introduction 5

Chapter 1: Sketching a Framework 7

1.1 Three Assumptions: Methodological Naturalism, Some Values and Consequentialism 7

1.2 Some Initial Definitions

1.2.1 Causal Determinism and Fatalism 12

1.2.2 Compatibilism, Incompatibilism, Libertarianism, Hard Determinism, Hard Incompatibilism, and Pessimism (or Impossibilism) 13

1.3 A More Precise Statement of My Main Theses 13

Chapter 2: A Variety of Freedoms and Responsibilities 16

2.1 Incompatibilist Theories of Free Will 16

2.2 Compatibilist Theories of Free Will 17

2.3 Theories of Moral Responsibility 27

Chapter 3: In Defence of Evaluative Compatibilism 37

3.1 Evaluating the Incompatibilist Freedoms 39

3.2 Evaluating the Compatibilist Freedoms 43

3.3 Evaluating the Incompatibilist Responsibilities 50

3.4 Evaluating the Compatibilist Responsibilities 54

3.5 Conclusion 59

Chapter 4: In Defence of Critical Evaluative Optimism 61

4.1 An Everyday Defence of Critical Evaluative Optimism 63

4.1.1 An Everyday Defence of the Evaluative Optimism Component 63

4.1.2 An Everyday Defence of the Critical Component 63

4.2 Four Challenges to Optimism about Free Will 64

4.2.1 The Challenge from Eliminative Materialism 64

4.2.2 The Challenge from Mechanism 68

4.2.3 The Challenge from Scientific Epiphenomenalism 69

4.2.4 The Challenge from Irrelevant Unconscious Influences and Confabulation 73

4.3 The Critical Component Is Becoming Less Critical 76

4.4 Conclusion 78

Conclusion 79

Bibliography 80

Introduction

Imagine that your choices and actions are causally determined by the laws of nature and by events that took place in the distant past, long before you were even born. In that case, it seems that your choices and actions are not free. After all, if you are not free with respect to the conditions that are causally sufficient for your choices and actions (the past and the laws of nature), then it seems that you are not free with respect to your choices and actions either (since they are the logical consequences of the past and the laws of nature), in the sense that you do not originate them: you are merely a link in an impersonal causal chain. Moreover, if it is physically impossible for you to make another choice or perform another action (because, given the past and the laws of nature, they *have* to happen), then, once again, it seems that your choices and actions are not free. Now imagine that your choices and actions are *not* causally determined: given the past and the laws of nature, it is possible to make more than one choice or to perform more than one action. In that case, too, it seems that your choices and actions cannot be free! After all, if more than one choice or action is consistent with the past and the laws of nature (if more than one choice or action is possible given all facts about the world and the agent in it), then it seems that which of them occurs is a mere matter of chance or luck. It is not *you* who determines or controls which of several choices you will make or which of several actions you will perform. Rather, it is a random process which is wholly outside of your control that forces you in one of several directions. In short, free will seems to require both that our choices and actions are undetermined *and* that they are not random; but that seems to be an impossible combination. That, in a nutshell, is the classical problem of free will.

Despite its intuitive appeal (it even convinced me – well, it affirmed my suspicion – that there is no such thing as free will when I first met it), it seems to me that it is a problem only for certain normative ethical theories – namely, for theories that state or presuppose that it would be a *bad thing* or a *moral tragedy* if we were to find out that we cannot do otherwise (in the sense mentioned above), that sufficient conditions for our choices and actions are laid down in the distant past, or that it is a mere matter of chance which of several choices we will make or which of several actions we will perform. And that means that it is *not* a problem for other theories. That, it seems to me, is something that it is very important to acknowledge and discuss. But, surprisingly perhaps, it hardly ever actually seems to be (explicitly?) recognized by participants in the free will debate. That is, they seem to believe that it is a problem *no matter what* normative ethical theory one endorses.

In this thesis, I will strive to counteract this, in my view, indefensible disposition, in the following way. In chapter 1, I will summarize and endorse both a particular kind of theory of what is valuable in itself (one that says that what is valuable in life is exhausted by positive and negative experiences and the perfection of our rational and moral capacities) and a particular kind of normative ethical theory (one that defines the rightness of everything in terms of their foreseeable consequences). In chapter 2, I will provide an outline of all the various freedoms that have (to my knowledge) been proposed as accounts of free will and of all the various responsibilities that have been proposed as theories of moral responsibility. After having performed that laborious task, I will use the distinctions that I have made in the first two chapters to argue, in chapter 3, for the first main thesis of this thesis: that the theory of value that I endorse has the important implication that all valuable freedoms are compatible with our actions and choices being determined by the past and the laws of nature – and with our actions and choices being probabilistically caused by the past and the laws of nature, too (although my emphasis will be on the former rather than the latter throughout this thesis); and that the theory of value that I will adopt entails, in conjunction with the normative ethical theory that I will endorse, that all valuable kinds of responsibility are compatible with our actions and choices being determined (probabilistically caused) by factors that are wholly beyond our control. Of course, challenges to the existence of valuable kinds of freedom and responsibility could come from corners other than determinism and indeterminism. In chapter 4, I will discuss four of the most widely discussed such challenges, which originate from philosophy, neuroscience and psychology; and I will conclude that none of them threatens the existence of any of the valuable freedoms and responsibilities, indicating, in conjunction with some everyday observations, that we have reason to be optimistic with respect to their existence. I will also argue, however, that everyday and scientific observations show that we have some of them to a lesser degree than would be ideal.

All in all, even if you do not agree (as is probable) with the particular theory of value and the particular normative ethical theory that happen to comport well with my ethical intuitions, and even if you (therefore) do not agree with any of my main theses, by the end of this thesis I at least hope to have convinced you of the importance of ethics to the free will debate. For some ethical theories, the problem of free will simply does not sting. That is something we should all acknowledge.

Chapter 1: Sketching a Framework

In this chapter, I will strive to accomplish three things. First, I will give an outline of, but (for reasons of space and relevance) will not extensively argue for, three positions that I will endorse throughout this thesis: that philosophy is continuous with science; a particular theory of what is valuable in life; and a particular normative ethical theory. Second, I will strive to explain some key concepts that I will need (or that I will presuppose) throughout this thesis. Finally, after having sketched this framework, I will be able to state my main theses more precisely.

1.1 Three Assumptions: Methodological Naturalism, Some Values and Consequentialism

Throughout this thesis I will, if doing so is relevant, make several assumptions. First, I will assume that ***methodological naturalism***, which is the thesis that philosophy is continuous with science, and that it therefore is rational to form one's metaphysical beliefs (one's beliefs about what the world is like), at least in part, on the basis of the scientific methods (Stoljar 2009, § 16), is correct. I will need this assumption in chapter 4, in which I will discuss four challenges that are widely thought to threaten the existence some valuable freedoms, three of which originate from neuroscience and/or psychology. After all, if philosophy would *not* be continuous with science, then it could be argued that scientific results cannot be relevant to the free will debate even in principle. It seems to me, however, that methodological naturalism is correct, for at least two reasons. First, the explanatory and predictive success of many sciences (especially the natural sciences) provides a *prima facie* good reason for thinking that they (or: the rigorous application of the scientific methods) succeed(s) in providing approximately true descriptions of the world. And, second, the persistent disagreement that plagues traditional metaphysics provides a *prima facie* good reason for thinking that it does not succeed (on its own) in providing (or at least establishing) approximately true descriptions of the world. After all, equally well informed and intelligent persons often persistently have conflicting intuitions about the same cases; and, in general, our *a priori* intuitions do not seem to have succeeded very well in leading us to the truth in the past. So it seems to me that one would be well advised not to trust *too* much on one's *a priori* metaphysical intuitions and to have faith in science instead.

Second, I will adopt a specific ***theory of the good***, that is, I will adopt a theory that specifies which things are *intrinsically* good or valuable, i.e., valuable "in themselves" (McNaughton 1998, §1), "for [their] own sake" (Weijers 2011, § 1.b) or "*non-instrumentally*" (Moore 2004, § 2), rather than valuable merely as a means for bringing about something that is valuable or preventing something that is disvaluable from coming into existence. The assumption of this theory will be needed in chapter 3 in particular, in which I will use it to argue that all valuable kinds of freedom (that have been proposed as accounts of free will) and responsibility are compatible with our choices and actions being causally determined by factors that are beyond our control; but I will also presuppose it in chapter 4, in order to argue that we have reason to be (critically) optimistic about the existence of those valuable kinds of freedom and responsibility. Before proceeding, however, I should note that I actually am a subjectivist about value, in the sense that I believe that nothing is *objectively* valuable, valuable independently of the subjective attitudes of individuals with regard to what is valuable: values do exist, but they exist not as properties of objects, but simply inhere in the experiences and judgments of subjects. That does not stop me from having intuitions, like everyone, about what is intrinsically valuable, however, and that is what I will base the endorsement of my theory on. In a nutshell, I believe (or feel, if you prefer) that three things have intrinsic value: pleasure (in a broad sense), the avoidance of pain (in a broad sense) and living a "virtuous" or "good" life (as opposed to a "pleasant" life). What do these consist in? On my account, an experience or feeling is pleasurable if and only if it is *positive* (good, valuable) in some way; and an experience or feeling is painful if and only if it is *negative* (bad, disvaluable) in some way. This is rather unenlightening, of course, because it does not specify at all which experiences or feelings are positive, and which are negative. Still, in many cases this is clear enough (for instance, eating tasty food, having a good idea and watching an excellent movie are (all things equal) positive experiences; and being assaulted or insulted, not being able to achieve what one wants to achieve and being forced to watch a horrible television series are (all things equal) negative experiences), so I shall restrict myself to that intuitive level. Moreover, I shall assume that one unit of pain is more disvaluable than one unit of pleasure is valuable, which implies that reducing pain is more important than increasing pleasure. (Imagine, for instance, as Weijers (2011, § 4.a) asks his readers to do, being given the chance to play the following game: if a coin that you are asked to toss lands on heads, then you are immediately overwhelmed by heavenly pleasure; but if the coin lands on tails, then you will suffer excruciating pain. Would you want to play the game? I certainly wouldn't.) Having discussed pleasure and pain, what does living a virtuous life consist in? That is a difficult question that I have given far too little thought, but for now I will take it to consist, in line with the perfectionist tradition (discussed by Brink 2007, 382, 391-392), in the exercise, development and expression of one's rational and practical (especially moral) capacities, that is, I will take it to consist in the perfection of one's receptivity for and reactivity to rational and moral considerations. How does the value of positive experiences and the avoidance of negative experiences compare to the value of the perfection of our responsiveness to reasons? Well, I'm not sure – my intuitions aren't clear about that, yet. Even so, for reasons of convenience, I will simply assume the default position – that they are of roughly equal importance.

Finally, I will adopt a rather specific form of ***consequentialism*** (hold your breath): expectable satisficing scalar direct global nondecisional personal impartial evaluative consequentialism.[1] I will need to assume this particular kind of normative ethical theory in chapter 3, in order to argue that all valuable kinds of moral responsibility are compatible with the determination of, say, our choices and actions by events in the distant past in conjunction with the laws of nature; and I will need it in chapter 4, when arguing that regular and scientific observations give us reason to be (critically) optimistic about the existence of all valuable freedoms (that have been proposed as accounts of free will) and responsibilities. Note, however, that I am a subjectivist about what moral rightness consists in as well: it certainly is true that certain things are right or wrong *given* one's beliefs, desires and values, but it is not the case, I believe, that certain things are right or wrong *no matter what* beliefs, desires and values one has.[2] So, what does the elaborate kind of consequentialism that I will endorse precisely consist in? Well, let's start with consequentialism. Consequentialism is a normative ethical theory (a theory of what it is right to do) that says that 'morality is *all* about producing the right kinds of overall consequences' (Haines 2006, introduction): it says that is right not to *honor* or *respect* value (i.e., 'to act on it or protect it at every opportunity' (Brink 2007, 383)), but to *promote* value (i.e., 'to take steps that lead to its greater realization overall' (ibid.)). Note, however, that consequences include not only that which some event, say, brings about, but also the event itself: its value matters too (Haines 2006, introduction). In other words, consequentialist moral theories 'make the moral assessment of alternatives depend in some way upon their value' (Brink 2007, 380), that is, they 'make the good explanatorily primary, explaining other moral notions (…) in terms of promoting value' (381). Deontological ethical theories, on the other hand, make the right primary, in that they regard certain kinds of actions as right (obligatory) or wrong (prohibited) on the basis of their intrinsic natures rather than based on the value of their consequences (McNaughton 1998, § 1; Brink 2007, 381); and virtue ethical theories make the idea of a virtuous character primary, in that they regard an action as right if it either is or would be performed by someone with a virtuous character (Brink 2007, 381). Having explained what consequentialism in general is, I will now explain each of the forms of consequentialism that I have introduced above. *Expectable* consequentialism states that whether something is right depends not on its actual consequences (as actual consequentialism asserts), but on its reasonably expectable consequences (the term comes from Haines (2006, § 1.d), but he uses it somewhat differently). If an action, say, has bad consequences, then it no doubt is *bad*; but if they were not foreseeable (e.g., when someone saves a drowning young Hitler), then it is

[1] I will probably not need *all* of those elements in this thesis. Still, they might have some important implications that I have overlooked, and for that reason I believe that it is prudent to describe all of them anyway.

[2] As one could also put it: certain moral beliefs can be true (in a “coherence” sense) *within* a particular moral framework; but they are not true (in a “correspondence” sense) *no matter what* moral framework you hold onto.

not *wrong* – or so it seems to me. (One could use a distinction between *objective* and *subjective* wrongness to the same effect. See Crisp & Chappell 1998, § 3 and Haines 2006, § 1.e.) *Satisficing* consequentialism states that whether something is right does not depend on whether its (expectable) consequences are *best* (as maximizing consequentialism says), but on whether its (expectable) consequences are *satisfactory* or *good enough* (above a certain threshold of value) (Brink 2007, 384-385; Sinnott-Armstrong 2011, § 6). How high should the threshold be? Well, it should be high enough to render it wrong (in our current circumstances) not to become a vegan and not to give a significant portion of one's income to charity; but it should be low enough so as not to make it wrong not to leave your avocations and loved ones in order to devote your life to saving strangers in desperate circumstances. That would certainly be a wonderful deed, but it isn't wrong not to do so. *Scalar* consequentialism states that something is more right (or less wrong) than some other thing if it has better (expectable) overall consequences than that thing and less right (or more wrong) if it has worse (expectable) overall consequences (Brink 2007, 383-384; Haines 2006, § 1.c, although they speak of "better" and "worse" rather than of more or less right and wrong). This implies that it is *more* right to devote one's life to helping many individuals in desperate circumstances, even though it isn't wrong not to do so. *Direct* consequentialism states that the normative properties of something (e.g., an action) depend only on the (expectable) consequences of that thing (rather than on the (expectable) consequences of something else, e.g., a rule or motive or character trait, as indirect consequentialism asserts) (Sinnott-Armstrong 2011, § 5). *Global* consequentialism states that it is true for *all* things (e.g., actions, rules, motives, character traits, practices and institutions) that their normative properties depend only on their (expectable) consequences (Hooker 2008, § 5; Sinnott-Armstrong 2011, § 5). In my view, this does not have to imply, say, that it is right both to develop a certain character trait (e.g., kindness) that is generally beneficial *and* to perform a certain action (e.g., torture) which is in tension with that character trait (Hooker 2008, § 5). Rather, it simply means that we should strive to find a sufficiently good mesh between all the different elements. *Nondecisional* consequentialism (the term is mine) states that direct (expectable) consequentialism should often not serve as a *guide* or *decision procedure*, but only as a *criterion* or *standard* of whether something is right or wrong (Brink 2007, 388; Hooker 2008, § 4; Sinnott-Armstrong 2011, § 4). So consequentialists should not, prior to every decision, strive to calculate which action produces a sufficiently good outcome. Why not? Well, that would have bad consequences, for several reasons. First of all, calculating prior to every decision which action would create sufficient value would often be counterproductive and inefficient: in many situations, we simply don't have time for doing such complex calculations; we often don't have adequate information about what consequences various acts that are open to us would produce; and the costs of obtaining such information will often not be proportional to the significance of the decision (Brink 2007, 387; Hooker 2008, § 4; McNaughton 1998, § 3). Second, even with adequate information, the calculations will often be quite complex, which means that mistakes will be made (Brink 2007, 387; Crisp & Chappell 1998, § 3; Hooker 2008, § 4; Sinnott-Armstrong 2011, § 4). Third, the calculations will quite often be distorted by the agent's biases (Brink 2007, 387; Crisp & Chappell 1998, § 3; Hooker 2008, § 4; McNaughton 1998, § 3): for instance, he may assign a disproportionate amount of value to his own interests and to those of people similar or close to him, or to near-term (rather than mid- or long-term) benefits. Finally, one may wonder whether anyone would want to have a relationship with someone who is continuously calculating the utility of several outcomes – including the value of the relationship (Haines 2006, § 3.d). If not, then doing so would be imprudent. So how *should* we make our decisions? Except in those unusual circumstances in which it *is* beneficial to calculate (even if only roughly) which of several actions would produce the best consequences, it is probably best to foster certain dispositions or character traits (e.g., kindness, prudence and truthfulness) and act in accordance with certain rules (e.g., “don't harm others”, “help others if needed” and “be honest”) that generally have good consequences in your circumstances (Brink 2007, 388; Crisp & Chappell 1998, § 3; Hooker 2008, § 4; McNaughton 1998, § 3; Sinnott-Armstrong 2011, § 4). This does not imply, however that direct (expectable) consequentialism simply is irrelevant to decision making. After all, a standard of what is right can still inform our deliberations about which of the available decision procedures, rules or dispositions we should endorse or foster; it can ask us to refine or modify our decision procedures (etc.) in the light of new knowledge and experiences, and given changed circumstances; and it can serve as a guide as to which rule (say) to adopt when several of the rules that we have adopted conflict with one another (Brink 2007, 388; Crisp & Chappell 1998, § 3; Sinnott-Armstrong 2011, § 4). Moreover, one often does not have to do any kind of complex calculations in order to see that one action will probably have better consequences than another. *Personal* or *sentient* consequentialism states that all things are valuable in virtue of the consequences that they have for the lives of sentient beings (Brink 2007, 382). *Impartial* or *agent-neutral* consequentialism states that 'whether some consequences are better than others does not depend on whether the consequences are evaluated from the perspective of the agent (as opposed to an observer)' (Sinnott-Armstrong, § 1). Two positions that I take to be strongly connected to this one are that *all* sentient beings count (“universal consequentialism”), and that they all count *equally* (are equally important: “equal consideration”), in the sense that (all things equal) one unit of value (one benefit) has as much value as a similar unit (a similar benefit), no matter who it gets to (Sinnott-Armstrong, § 1). Finally, *evaluative* consequentialism states, rather trivially perhaps, that whether something is right depends only on the *value* of the (expectable) consequences of that thing (rather than on non-evaluative features of the consequences) (Sinnot-Armstrong, § 1).

1.2 Some Initial Definitions

Having provided an outline of the various positions that I will endorse throughout this thesis, I will now explain several key concepts that are essential for understanding my main theses and arguments.

1.2.1 Causal Determinism and Fatalism

First, *determinism* roughly is the thesis that for every event 'there are conditions (…) whose joint occurrence is (logically) sufficient for [their] occurrence' (Kane 2011, 4): given those conditions, the events inevitably *have* to occur. So if our universe were to be deterministic, and if it were "rolled back" (for whatever reason) an indefinite number of times to some moment in the distant past, then it would proceed in *exactly* the same way every single time. *Causal* determinism, the thesis on which I will direct my attention[3], is one species of determinism, and roughly is the thesis 'that the course of the future is entirely determined by the conjunction of the past and the laws of nature' (Timpe 2006, § 3.a). So, if causal determinism is true, then every event (except the first one, if there is a first one), including our choices and actions, is *causally* determined or fixed by *antecedent* events together with the laws of nature. More neutrally (without mentioning events and determination), 'causal determinism is the doctrine that a complete statement of the laws of nature and a complete description of the temporally nonrelational or "genuine" facts [facts that do not refer to other times] about the world at some time T *entail* every truth about the world after T' (Fischer 2007, 322). Furthermore, the truth of causal determinism requires that every event has a determinate state or description at any given time (Hoefer 2010, § 2.5), and (ii) that the laws of nature apply to everything in the universe (Hoefer 2010, § 2.5; Vihvelin 2011, § 1). Determinism figures prominently in my first main thesis, so it's very important that you understand what it consists in.

At the outset, I would like to distinguish (causal) determinism from another thesis, fatalism, with which it sometimes is conflated. *Fatalism* is the thesis that it is *one's fate* to have a certain future (or to enjoy or suffer a certain event or series of events), and that one's psychological processes (one's beliefs, desires, choices, decisions, etc.) and actions simply do not make a difference as to whether that particular future comes into being (Eshleman 2009, § 1). So if fatalism is true, then one's future (or part of it) is not the result of processes that go *through* oneself but the result of processes that go *around* oneself: one's mental states and events and one's actions are *bypassed*, because certain events simply *had* to happen even if earlier events would have been different (Nahmias 2011, 561-562). Determinism, however, does not entail any such thing: if determinism is true, then one's mental states and events and one's actions (probably) *do* make a difference and *are* causally relevant as

[3] Even so, in what follows I will often leave away the "causal". When doing so, I will mean *causal* determinism.

to whether a particular future is realized, in the sense that its realization is (in part) dependent upon which choices and decisions one makes, which actions one performs, and so on..

1.2.2 Compatibilism, Incompatibilism, Libertarianism, Hard Determinism, Hard Incompatibilism, and Pessimism (or Impossibilism)

Compatibilism, as I use the term, is the thesis that free will and moral responsibility are compatible with determinism (McKenna 2011, 176), i.e., that it is possible both to possess free will and to be morally responsible and to be determined in all one's choices and actions (Timpe, 2006, § 3c). *Incompatibilism* is the thesis that free will and moral responsibility are incompatible with determinism, i.e., that it is impossible both to possess free will and to be morally responsible and to be determined in all one's choices and actions. *Libertarianism* is the conjunction of incompatibilism and the thesis that at least some of us, at times, possess free will and are morally responsible (Haji 2009, 19). *Hard determinism* is the conjunction of incompatibilism and the thesis that determinism is true, which implies that free will and moral responsibility do not exist. *Hard incompatibilism* is the thesis that free will and moral responsibility are not compatible with determinism, but that they are not compatible with (event-causal) indeterminism either. *Pessimism* or *impossibilism* is the thesis that the existence of free will and moral responsibility is impossible. And *optimism* is the thesis that the existence of free will and moral responsibility is not only possible, but actual, i.e., it is the thesis that free will and moral responsibility exist in the actual world – in our world.

1.3 A More Precise Statement of My Main Theses

Having explained the main ethical positions that I will endorse and the key concepts that I will use (or presuppose) throughout this thesis, I am now able to state my main theses more precisely. My first main thesis is that *compatibilism* is correct, i.e., that free will and moral responsibility are compatible with determinism. More precisely, I will argue in favor of a particular *version* of compatibilism, which I will call "*evaluative* compatibilism". This is the thesis that, of all the freedoms that have been proposed as (parts of) accounts of free will, all *valuable* freedoms are compatible with determinism (the freedom component); and that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all *valuable* responsibilities are compatible with determinism (the responsibility component). This kind of compatibilism should be distinguished in particular from what I will call "*verbal* compatibilism", which states that what "free will" and "moral responsibility" mean is compatible with what "determinism" means. In my opinion, we should not direct our attention at this sort of compatibilism, for two reasons. First, it seems to me that the dispute over whether what “free will” and “moral responsibility” mean is compatible with what “determinism” means simply is *pointless*, because there is *no fact of the matter* as to what “free will” and “moral responsibility” mean. After all, all that *could* make it the case that “free will” and “moral responsibility” have a certain meaning in our linguistic community is widespread intersubjective agreement that they have those meanings (or so it seems to me). But given that philosophers have offered wildly different accounts of what those terms refer to and do not even agree (at least in the case of free will) on their minimal meaning, such agreement seems to be absent. So “free will” and “moral responsibility” do not seem to have 'one single and clear meaning in ordinary language' (Kane (2011, 35, note 23) says that Richard Double has given this argument with respect to “free will”): rather, the public language that we speak (including philosophical language) seems to contain several different concepts of “free will” and “moral responsibility”, none of which is *the* meaning of the terms.[4] Even if I am wrong about this, however, there is a second reason why we should not focus on verbal com-patibilism – because it is entirely *uninteresting*. Given that it is a purely semantic issue what “free will” and “moral responsibility” mean, it is hard to see what important fact could possibly depend on it. Suppose that experimental philosophers discover tomorrow that “free will” and “moral responsibility” have a certain meaning in ordinary language that is incompatible with the meaning of “determinism”. It seems rather implausible to suppose that compatibilists would simply throw in the towel after having heard the news. Rather, they would probably simply modify their vocabulary and argue that even if “free will” (say) is not compatible with determinism, some other freedom *is* – and that, they will argue, is what we should care about, not that silly “free will”.

My second main thesis, which I will call “critical evaluative optimism”, states that of all the kinds of freedom that have been proposed as (parts of) accounts of free will, all valuable freedoms exist in the actual world, and that of all the kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities exist in the actual world (the evaluative optimism component); but that scientific and everyday observations do show that we have some freedoms and responsibilities to a lesser degree than would be desirable (the critical component).

So those are my main theses! Having explained those, I am now also able to state more precisely what I will do in the remaining parts of this thesis. In the next chapter, I will provide an outline of the most prominent kinds, at least, of incompatibilist freedom that have been proposed as accounts of free will; of the most prominent kinds of compatibilist freedom that have been proposed as ac-

[4] Besides Richard Double's, Kane (2011, 26-27) describes similar arguments that have been given by Ted Honderich (who has argued that we have a conception of more than one kind of freedom, i.e., that 'freedom has a number of different meanings' (27)) and Saul Smilansky (who has argued that we should recognize that some important freedoms are incompatible with determinism, but that some other important ones are compatible with it). A similar sentiment (though not with respect to *ordinary* language) is echoed by Galen Strawson, when he says that '[m]ore than 200 senses of [“free will”] have been distinguished' (Strawson 1998, § 1) by philosophers. Finally, Jesse Prinz (Center for Inquiry – New York (2011)) says that he believes that we have many different notions of freedom.

counts of free will; and of the most prominent kinds of responsibility that have been proposed as accounts of moral responsibility. In chapter 3, then, I will evaluate all those kinds of freedom and responsibility based on the theory of value and the normative ethical theory that I endorse, in order to argue that all kinds of freedom that are incompatible with determinism are without value; that some of the kinds of freedom that are compatible with determinism are valuable; and that none of the kinds of responsibility that are incompatible with determinism are valuable, but that some of the kinds of responsibility that are compatible with it are. If this is correct, then, evidently, evaluative compatibilism would be on firm ground. Finally, in chapter 4, I will discuss four challenges that originate from philosophy, neuroscience and/or psychology that have been widely though to threaten some valuable freedoms and responsibilities. I will argue that neither of these challenges threatens the *existence* of any of the freedoms and responsibilities that I have concluded to be valuable; but that one of these challenges (not to mention loads of everyday observations) does (do) threaten the *degree* to which we have some of the valuable kinds of freedom. Once again, if this is true, then critical evaluative optimism would seem to be a position that is well-founded.

Chapter 2: A Variety of Freedoms and Responsibilities

In this chapter, I will strive to provide an overview of the most prominent kinds of freedom that have been proposed as accounts of free will and of the most prominent kinds of responsibility that have been proposed as accounts of moral responsibility. In section 2.1, I will summarize the most prominent kinds of *incompatibilist* freedom that have been proposed as (parts of) accounts of free will. In section 2.2, I will provide an outline of the most prominent kinds of *compatibilist* freedom that have been proposed as (parts of) accounts of free will. And, in section 2.3, I will provide an overview of the most prominent kinds of responsibilities that have been proposed as (parts of) accounts of moral responsibility. As I've explained above, all this is necessary for defending both of my main theses – one cannot argue that all valuable kinds of freedom and responsibility exist and are compatible with determinism without having accounts of those things in place that can be evaluated.

2.1 Incompatibilist Theories of Free Will

There are (at least) three prominent incompatibilist kinds of freedom that have been proposed as (parts of) accounts of free will. First, there are what I will call ***absolutist*** theories of freedom, according to which true freedom is *absolute*. One theory of this sorts says that free will consists in intentionally bringing it about that one is the way one is in certain crucial mental respects, i.e., that one has a certain mental nature, or a certain character (Strawson 1998, § 3)[5]. Another theory within this framework states that free will consists in being an uncaused cause of oneself, that is, it consists in causing oneself without being 'causally influenced by external causal factors, including [one's] own character' (O'Connor, T., 2010, § 3.2)[6]. This, it seems to me, is captured well by what Knobe and Nichols (2011, 539-540) call "the executive conception of the self", which is the view that 'psychological states do not constitute the self, they *belong* to the self, and the self makes its decisions *in light of* the psychological states, but not as a simple consequence of these states' (540), so that 'the self is an executive that stands apart from the particular mental states that inform her decision' (ibid.). It differs from the first theory because that theory does *not* say that one may not be causally influenced by external causal factors. Rather, it only states that true freedom requires that one is not (or perhaps not *significantly*) causally influenced by causal factors that one has not intentionally brought into existence: if one *has* intentionally brought it about that one has a certain character, say, then being influenced by it does not diminish one's freedom. But both theories are incompatibilist: the first theory is incompatibilist, because the truth of determinism implies that one cannot intentio-

[5] This conception of free will has been defended by Galen Strawson in particular. I should note, however, that he actually believes that it is *impossible* for anyone to possess this kind of freedom.

[6] According to O'Connor (2010, § 3.2), this conception of free will has been defended by, among other people, C.A. Campbell, Jean-Paul Sartre and Duns Scotus,

nally originate one's mental nature in the sense required (from scratch); and the second, because not being causally influenced by external causal factors evidently is inconsistent with determinism.

Second, there are theories of free will that insist that free actions are events that are ***uncaused***. These theories do not explain the freedom of actions in terms of what they are caused by or in terms of how they are caused (what internal causal structure they have). Rather, their freedom is explained in terms of the reasons or intentions or purposes of their agent, in terms of what they are aimed or directed at. (See "noncausal theories of free will" below for a more extensive discussion of these noncausal sorts of requirements.) Causal determinism requires that every event has a (and is fully determined by its) cause, so this kind of freedom is not compatible with causal determinism.

And, third, there is ***the libertarian account of the ability to do otherwise***. Libertarians argue that, in a world w, one is able, at time t, do other than some action A at time t* (which may be identical to or later than t) only if there 'is a possible world with (i) exactly the same laws [of nature] as w until t, and (ii) exactly the same past right up to, or just prior to t, as w, in which S, at t, does something other than A at t*' (Haji 2009, 36). More intuitively, a person can do otherwise only if she can do other than she does *given* the past and *holding fixed* the laws of nature: when a person is using her free will, 'what she is doing is selecting from a range of different options for the future, each of which is possible given the past and the laws of nature' (Timpe 2006, § 4a). It should, though, be noted, that, for libertarians, the ability to do otherwise is only a starting point, in that, for them, the ability to do otherwise is a *necessary* but not a sufficient condition for free will. Some of them, for instance, also require "executive power"; others require "agent causation"; still others require one or more compatibilist conditions; and yet others impose no causal requirements at all (for the first requirement, see above; for the other requirements, see section 2.2 below). Excepting the last group, libertarians either locate the required indeterminism in 'the immediate causal antecedents' (Clarke 2008, § 2.2) of a free action, or earlier in the deliberative process, such as in the 'coming-to-mind of considerations that bear on choice or in the formation of preferences' (Kane 2011, 22).[7]

2.2 Compatibilist Theories of Free Will

To my knowledge, there are (at least) ten prominent compatibilist kinds of freedom (freedoms that are compatible with determinism) that have been proposed as (parts of) accounts of free will. First, there are ***agent-causal*** theories of free will. These say that an agent is a persisting substance, and that a mental event is (directly: its freedom is not derived from the freedom of other mental events

[7] According to Clarke (2008, § 2.1, § 2.2), the event-causal view of this type (i.e., the view that it is a necessary condition of free actions that they are indeterministically caused by certain events involving or within the agent) has been defended by, among others, Mark Balaguer, Laura Ekstrom, Robert Kane, Alfred Mele and Robert Nozick. This is confirmed in many of the other texts that I have read. I mention other prominent libertarians elsewhere.

(Haji 2009, 43)) free if and only if it is caused by the *agent*. This, in turn, requires that the agent possesses a special kind of causal power, an ontologically primitive (and distinctively personal (O'Connor 2010, § 3.2)) power of causation by the agent (substance), which does not consist in and cannot be reduced to ordinary sorts of causation by states or events within or involving the agent (Clarke 2008, § 3; Haji 2009, 43-46; Kane 2011, 20; O'Connor 2010, § 3.2).[8] As far as I know, all proponents of (the coherence of) agent-causation (as an account of free will) endorse the principle that free will requires that neither the agent's causing an event nor what the agent causes may be determined by prior causes: only *then*, then argue, can the agent be the ultimate source that free will requires. Still, it seems to me that compatibilists *could* endorse agent-causation as well: after all, they could argue that free will requires that we are agent-causes, without arguing that it requires that both our agent-causing and what we're causing are not deterministically caused by prior events. It could be objected that since a substance 'is not the kind of thing that can itself be an effect' (Clarke 2008, § 3; Haji 2009, 45; both of them give the exact same sentence), the causation by a substance cannot be deterministically (or otherwise) caused even in principle, and perhaps this is true. But that leaves open the possibility, as Haji (2009, 46, 180) says, that free actions are caused by the agent *and* deterministically caused by events or states within or involving the agent. So, given that all that is required for a freedom to be compatibilist is that it is *possible* for it to exist in a deterministic universe, agent-causation, on itself, seems to be a compatibilist freedom.

Second, there are ***noncausal*** theories of free will. These impose no causal requirements on free actions at all, that is, they 'require neither that a free action be caused by anything nor that it have any internal causal structure' (Clarke 2008, § 1). According to Clarke (2008, § 1), they generally argue that all that free action requires is that 'is or begins with a basic mental action' (ibid.) (in contrast to a nonbasic action, an overt bodily action, which is caused by a basic mental action), a decision or choice, which is an intrinsically active doing, intrinsically something that I do, and has an intrinsic purposiveness or intentionality, in that 'when one makes a decision, intrinsic to the decision is one's intending to make that very decision' (ibid.). O'Connor (2010, § 3.2) puts it similarly: according to him, noncausal theories state that 'I control my volition or choice simply in virtue of its being mine – its occurring in me. I do not exert any kind of special causality in bringing it about; instead, it is an intrinsically active event, intrinsically something that I do'. So the freedom of actions is, on this view, not to be explained in terms of an agent's freely *causing* or *bringing about* something (in terms of a relation between an agent and an action (Pink 2010, 305)), but in terms of the *reasons* or *intentions* or *purposes* an agent has, in terms of his striving to attain a certain end or goal (Clarke 2008,

[8] According to Clarke (2008, § 3), this view (that it captures what free will is) has recently been defended by Roderick Chisholm, Randolph Clarke, Timothy O'Connor, Derk Pereboom and Richard Taylor, among other philosophers.

§1; O'Connor 2010, § 3.2; Pink 2010, 305-306).[9] As with agent-causal freedom, to my (rather limited, to be sure) knowledge, all proponent of noncausal theories of free will are libertarians. Still, compatibilists could easily adopt a noncausal account of free will as well, because noncausal theories, given that they impose no causal requirements on free actions, simply do not entail that a free action must be uncaused, nor that it may not be deterministically caused.

Third, there is the conception of free will as ***active control***. This conception of free will is based on the so-called event-causal theory of action, which states that an action is a particular sort of event – one that is (nondeviantly) caused by certain mental events, in particular motivational mental states (e.g., desires) and cognitive mental states (e.g., beliefs), that occur within an agent. In line with this theory, one exercises active control when certain mental events within oneself are the (nondeviant) proximate cause of certain events involving oneself; and one acts for a reason if one's having that reason is the cause of relevant action (Haji 2009, 38-40).[10][11] Because determinism does not imply that certain kinds of mental states or events are never (cannot be) the (nondeviant) proximate cause of events involving ourself, this freedom obviously is compatible with determinism.

Fourth, it has been argued that free will consists in ***the unencumbered ability to do what one wants to do*** (McKenna 2009, § 3.1). All that free action requires, according to this theory, is '(i) to have the *power* or *ability* to do what you will (desire or choose or try) to do, and this entails (ii) an absence of [internal, such as paralysis or mental impairment, and external, such as being physically restrained or coerced] *constraints* or *impediments* preventing you from doing what you will (desire or choose or try) to do' (Kane 2011, 11). In short, it requires that we have 'the *power* (which equals *ability* plus *opportunity*) to do what [we] will to do' (ibid.; see also McKenna 2009, § 3; O'Connor 2010, § 1.1; and Timpe 2006, § 1).[12] Determinism does not imply that it is impossible have the power (ability plus opportunity) to do what one wants to do, so this kind of freedom is compatible with it.

A fifth branch of theories consists of ***compatibilist conceptions of the ability to do otherwise***. I have encountered five theories of this sort. First, some compatibilists (the “classical compatibilists” (McKenna 2009a, § 3; Kane 2011, 12)) conceive of the ability to do otherwise as *the ability to do otherwise if one would want to do so*. These compatibilist build on the view of free will as the unim-

[9] According to Clarke (2008, § 1), this theory of free will has recently been defended by Carl Ginet, Hugh McCann and Thomas Pink, among others.

[10] Although most (not all) theorists in the field consider active control to be a *necessary* condition of free action, to my knowledge no philosopher has argued that active control is by itself *sufficient* for free action.

[11] Hoefer (2010, § 2.1) argues that 'neither philosophers' nor laymen's conceptions of *events* have any correlate in any modern physical theory. The same goes for the notions of *cause* and *sufficient cause*' (note omitted). If this is true, then perhaps active control is nonexistent for that reason. But it seems plausible, at least, that conceptions of events and causes can be formulated that are existentially equivalent to common sense ones, and *do* obtain. To be honest, though, I literally know nothing about this subject matter, so, for the rest of the thesis, I will leave it to one side.

[12] According to Kane (2011, 12), this conception of free will has been defended by, among other people (it used to be very popular), A.J. Ayer, Donald Davidson, Thomas Hobbes, David Hume and Moritz Schlick.

peded ability to do what one wants to do in order to offer a *conditional* or *hypothetical* analysis of the ability to do otherwise, according to which one is able to do otherwise if one *would* do otherwise (no constraints or impediments would prevent one from doing so), if one would want (desire, choose, or try) to do so (Kane 2011, 12). And this, the classical compatibilist holds, importantly and 'accurately distinguishes those actions [an agent] would have performed if she wanted [i.e. that were within her power, or within her control], from those actions she could not have performed even if she wanted' (McKenna 2009a, § 3.3). Since this is an analysis only of what an agent would have done under certain *counterfactual* conditions (i.e., of what she would have done if her character or psychological states would have been different), this is consistent with her being causally determined to do what she does under *actual* conditions (Haji 2009, 34-35; Kane 2011, 11-12; McKenna 2009, § 3; O'Connor 2010, § 1.1; Timpe 2006, § 1).[13] Second, in recent times, some compatibilists have provided an analysis of the ability to do otherwise that builds on an intuition that is similar to the one that motivated the classical compatibilists: an agent can do otherwise if she has the *ability* or *capacity* (that she is determined not to *exercise*) to do otherwise. More precisely, they analyze the ability to do otherwise in terms of a *bundle of dispositions* in the following way: first, they 'hold fixed the relevant causal base or underlying structure of an agent's disposition' (McKenna 2009a, § 5.1.5) to perform some action; second, they 'consider various counterfactual conditions in which that causal base or underlying structure operates unimpaired' (ibid.); finally, if, they argue, the agent performs (or would perform) the action in a wide enough range of counterfactual conditions (if she would perform it in the right set of possible worlds (Levy & McKenna 2009, 112)), then, even though she is determined not to perform it, she does have the *ability* to perform the action (Levy & McKenna 2009, 112; McKenna 2009a, § 5.1.5).[14] As with the classical compatibilist analysis, since the conception of the ability to do otherwise as a bundle of dispositions only mentions *counterfactual* conditions, it is not threatened by determinism. Third, there is the view that the ability to do otherwise (more specifically, “avoidability”) consists in *having evolved to do otherwise*, that is, it consists in having evolved (via a process of natural selection) in such a way that one is now able to avoid certain outcomes (e.g., diseases, or starvation) that previous members of one's species (or of a species that one's species derives from) could not avoid and to realize certain others (e.g., complex moral behavior, or linguistic communication) that previous members of one's species could not realize (McKenna 2009a, § 5.2.3).[15] Causal determinism does not rule out that species can evolve in such a way that they are now able to avoid outcomes that they could not avoid before, so this kind of free-

[13] According to Kane (2011, 12), this conception of the ability to do otherwise has been defended by, among other people, A.J. Ayer, Donald Davidson, David Hume and Moritz Schlick (not Thomas Hobbes (McKenna 2009, § 3.3)).

[14] According to McKenna (2009a, § 5.1.5), this analysis of the ability to otherwise has been defended by Michael Fara, Michael Smith and Kadri Vihvelin.

[15] According to McKenna (2009a, § 5.2.3), this view has been defended by Daniel Dennett.

dom is compatible with determinism. Fourth, there is a conception of the ability to do otherwise as *being able to do otherwise in similar circumstances*, that is, to cite what is (as far as I know) its only proponent, it consist in the ability 'to exhibit *adaptive flexibility* while acting for intelligible reasons which are authentic (e.g., originate from one's own intentions, norms and values)' (Walter 2011, 517, italics SG). Since determinism is inconsistent only with the ability to do otherwise in the exact *same* circumstances, it is not inconsistent with this sort of flexibility. Finally, there is the view of "unavoidability" (or the *in*ability to do otherwise) as the *bypassing of* (or "going around") *a person's agency* (i.e., it is causally irrelevant as to whether a particular outcome is realized). As McKenna (2009a, § 5.1.3) says, on this view, 'when we say that such and such is unavoidable for a person, we have in mind "selective" contexts in which the facts pertaining to the unavoidability have nothing to do with that person – the facts bypass that person's agency altogether'.[16] As a mirror image of this view, the ability to do otherwise might perhaps be construed, similar to the classical compatibilist conception of this ability, as acting without having one's agency bypassed. Determinism does not entail that one's agency is always bypassed, so it is compatible with this ability.

A sixth theory states that it is a necessary condition for one's action being free that one has ***taken responsibility for the source of one's action***. More precisely, the freedom of an action requires that it is caused by a mechanism that the agent has come to own 'by means of a process whereby she takes responsibility for the mechanisms giving rise to her action' (McKenna 2009b, § B.2). This, in turn, requires (i) that the agent sees herself *as an agent* (McKenna 2011, 191), that is, as someone who is able to influence the course of events around her in virtue of being able to deliberate, choose and act, (ii) that she believes that it is apt or appropriate or proper that members of the moral community place moral expectations on her and direct their reactive attitudes at her, (iii) that the beliefs specified in (i) and (ii) are appropriately based on her evidence, and (iv) that she has acquired these beliefs 'through an appropriate means' (McKenna 2009b, § B.2) (which is taken to exclude freedom-undermining sorts of manipulation, such as having one's brain tampered with or brain-washing) (Haji 2009, 48-19; McKenna 2009b, § B.2; McKenna 2011, 191).[17] Given that determinism does not imply that none of these conditions can obtain (but only that *whether* they obtain is fixed by the past and the laws of nature), the kind of freedom is compatible with determinism.

Seventh, there are theories that explain free will in terms of a kind of ***harmony*** or ***mesh*** between certain kinds of structural mental aspects or elements of agents. They are sometimes also called "content-neutral" or "procedural" (Dryden 2010, § 2), because they deny that free will requires that

[16] According to McKenna (2009a, § 5.1.3), this conception of "unavoidability" has been defended by Michael Slote – and it has been endorsed by Daniel Dennett, among other people.

[17] According to McKenna (2009b, § B.2), this view has been defended as necessary (not sufficient) for free will (more precisely, the kind of freedom that is required for moral responsibility, as they define it) by John Martin Fischer and Mark Ravizza. For another necessary condition (responsiveness to reasons), see below.

the agent has to endorse certain values or principles, or has to perform certain kinds of actions, or “coherentist” (Buss 2008, § 2), because it is not the *content* or the *origin* or the *force* of, say, a person's desires that render them free (or unfree): rather, 'an agent governs her own action if and only if her motives (the desires that move her to act) cohere with (are in harmony with) some other attitude that represents her point of view on the action' (ibid.), i.e., if she *endorses* them or *identifies* with them in some way. There are at least five theories of this sort. First, there is the theory that free will consists in *being moved to action by the desires that one identifies with*. Before being able to state this more, precisely, however, I need to introduce some concepts: “first-order desires” are desires that have actions as their objects; “higher-order desires” are desires that have lower-order desires as their objects; an agent's “will” consists in the having of a first-order desire that either is or would be effective (absent constraints or impediments) in moving her all the way to action; and an agent's “volition” consists in the having of a higher-order desire that a certain first-order desire will be effective, i.e., that it will be one's will, so that one will be moved to action by it (McKenna 2011, 178, 197). Now, an agent's *will* is free if it consists in a first-order desire that an agent *actively* identifies with (i.e., *she* determines her own will (McKenna 2009b, § A)). This means that she has an *unopposed* volition (i.e., it is not in tension with another volition at the same level and it is the highest-order volition that the agent has), with which she is *satisfied*, in the sense that she has no inclination to change it, to be moved to action by that first-order desire (McKenna 2001, 178-179, 182). And an agent's *action* is free if it is brought about by her will (i.e., if her will is successful in moving her to action) (178). So, in short, a person 'acts of her own free will if and only if her action issues from the will she wants' (McKenna 2009a, § 5.3.2), or, in other words, if she acts on the desires that she wants to act upon, that is, if she acts on the desires that she identifies with (Buss 2008, § 2; Dryden 2010, 2.a.i; Ekstrom 2010, 101-104; Fischer 2010, 212-213; Haji 2009, 46-47; Kane 2011, 13-14; Levy & McKenna 2009, 108; McKenna 2009a, § 5.3; McKenna 2009b, § A; McKenna 2011, 178-184; O'Connor 2010, § 3; Strawson 1998, § 1).[18] Since nothing in determinism implies that one cannot be moved to action by desires that one wants (in the sense specified above) to be effective, this freedom is compatible with determinism. Second, it has been argued that free will consists in *being moved to action by desires that are harmonious with one's values*. More precisely, conceive of an agent as having within herself (among other things) two sorts of systems: a “valuational” system, which consists in what an agent considers to be 'valuable, good, or desirable' (McKenna 2011, 184), or 'worth pursuing' (Kane 2011, 14) (not in the sense that she *judges* some things to be (objectively)

[18] According to McKenna (2009a, § 5.3) (and everyone else, really), this view has been defended by Harry Frankfurt. It should, however, be noted that it only accounts for the kind of freedom of will and action that is, in his view, required for moral responsibility. It does *not* account for the ability to do or will *otherwise*, for the ability 'to make some other first-order desire [one's] will' (McKenna 2011, 179). According to McKenna (2011, 179) and O'Connor (2010, § 2), Frankfurt remains uncommitted with respect to whether this freedom is compatible with determinism.

valuable, but in the sense that she *values* them or *cares* about them (McKenna 2011, 197-198, note 13)) and originates in the agent's reason, and a “motivational” system, which consists in what an agent *desires*, or what she is motivated by. Now, an agent acts freely, on a particular occasion, when her motivational system 'works in harmony with' (185) her valuational system (Buss 2008, § 2; Kane 2011, 14; McKenna 2009b, § A; McKenna 2011, 184-185; O'Connor 2010, § 1.2).[19] Determinism (obviously) does not entail that one's values can never be harmonious with one's motives, so this kind of freedom, too, is compatible with determinism. A third kind of “mesh theory” says that free will consists in *being moved to action by desires that “fit into” one's larger plans*. More precisely, an action is free if it is *intentional*; and an action is intentional if it is 'embedded in' (McKenna 2011, 186) the larger ('more or less long-term' (Buss 2008, § 2)) plans or policies that govern one's life, which are plans that (i) make possible the bringing into practice of one's other plans, and (ii) 'govern (by constraining or fostering)' (ibid.) other plans that one might come up with (Buss 2008, § 2; Kane 2011, 14-15; McKenna 2009b, § A; McKenna 2011, 186).[20] Once again, determinism does not exclude the ability to act on desires that properly mesh with one's larger plans, so this kind of freedom is compatible with determinism. Fourth, there is a “mesh theory” that says that free will consists in *being moved to action by preferences that are coherent with one's other preferences*. More precisely, understand a preference as holding 'a certain first level desire to be good' (Dryden 2010, 2.a.iii) on the basis of a process of 'critical evaluation with respect to the agent's conception of the good' (Ekstrom 2010, 104). Now, an agent acts freely if she is nondeviantly moved to action by preferences that have been formed and are maintained 'without the coercive influence of another agent which the agent herself has not autonomously arranged' (ibid.) and that cohere 'with the agent's convictions and other reflectively evaluated desires' (ibid.) (Dryden 2010, 2.a.iii; Ekstrom 2010, 104-105; Kane 2011, 23).[21] Evidently, causal determinism is compatible with being moved to action by preferences that are mutually supportive as well. Finally, some compatibilists have argued that free will consists in *performing an action that can properly be attributed to oneself*. On this conception of free will, acting freely does not require that an agent performs the action *voluntarily* (that she has control over whether she performs the action or not), but only that it can properly be *attributed* to her, in the sense that 'it is expressive of her attitudes and values' (Levy & McKenna 2009, 115). And

[19] According to McKenna (2009b, § A), this view has been defended by Gary Watson.

[20] According to McKenna (2011, 186), this view has been defended by Michael Bratman. He does, however, note (198, note 16) that Bratman does not intend his theory to be an account of *free will* or even the kind of freedom that is required for moral responsibility. Rather, it is an account of what *personal autonomy* consists in. Still, it can usefully be (and has been) interpreted as a conception of free will, and that is what I will do as well.

[21] This view has been defended by Laura Ekstrom (2010, 104-105). Although the above certainly is compatible with determinism, she is, in fact, a libertarian, in that, for her, the above is only *part* of the story. More precisely, she requires (according to Kane (2011, 23)) that one's preferences are caused *but not determined* by the considerations that inform the formation of one's preferences; and this requires that determinism is false.

an action is expressive of an agent in this sense if it expresses her *judgment-sensitive* attitudes, which are attitudes that are sensitive to reasons in *ideally rational* agents (i.e., attitudes that ideally rational agents would only have when 'they judge there to be sufficient reason for them' (ibid.)). (See also McKenna 2009b, § G.1).[22] Yet again, determinism by no means implies that an action cannot be expressive of one's judgment-sensitive attitudes, so this freedom is compatible with it.

Eighth, there is a conception of free will as consisting in ***self-control***. More precisely, it conceives of free will as the ability to do what one judges it best (or better) to do in the face of a strong desire not to perform that action (McKenna 2009b, § F).[23] Or, as Walter (2011, 523) puts it (while discussing the neuroscientific concept of "volition"), 'it can be defined as a set of self-regulatory functions that enables subjects to realize their chosen intentions in an adaptive way, based on anticipated future outcomes, and to pursue their long-term goals in the face of conflict, temptations, and distractions'. In other words, it consists mainly in the trait of being able to translate the desires that one identifies with, that are harmonious with one's values and that "fit into" one's larger plans into action *over time* or *in the long run*, even when it is difficult to do so. Determinism evidently does not entail that we can never exercise self-control, so this kind of freedom is compatible with it.

Ninth, free will can be conceived as ***responsiveness to reasoning***. This consists, in the words of Buss (2008, § 2), in 'the capacity to evaluate one's motives on the basis of whatever else one believes and desires, and to adjust these motives in response to one's evaluations. It is the capacity to discern what "follows from" one's beliefs and desires, and to act accordingly'. In other words, responsiveness to reasoning consists in the ability to *recognize* what reasons one's beliefs and desires give one, and to *react* in an appropriate way (to change one's mind, or to act differently than before) in response to the recognition of these reasons. It certainly is content-neutral with respect to the *content* of one's beliefs and desires, in the sense that it does not require that one holds certain kinds of beliefs or makes certain kinds of evaluative judgments and has certain kinds of desires, and it is also (by itself) content-neutral with respect to the *origin* of one's beliefs and desires. But it does appeal to some standards or principles insofar as it positively recommends that agents engage in the critical evaluation of their beliefs and desires, and insofar as it presupposes that reasoning must satisfy certain stan-

[22] According to Levy & McKenna (2009, 115), this view has been defended by T.M. Scanlon and Angela Smith. It should be noted, however, that, although both of them consider this condition to be sufficient for *being* morally responsible for an action, they (i) do not understand this condition as a *freedom* condition, and (ii) they argue that *holding* responsible and blame (and the imposition of sanctions) *do* require the satisfaction of freedom (or control) conditions. Inversely, George Sher argues that moral responsibility requires control because it a causal notion, but that we can be blameworthy (at least) for traits for which we are not responsible and (perhaps) for actions which we do not control because blame is *not* a causal notion (Levy & McKenna 2009, 128).

[23] According to McKenna (2009b, § F), this view has been defended by Alfred Mele. He is not a compatibilist per se, nor is he a libertarian. Rather, he is an *agnostic* about whether or not free will is compatible with determinism, who has formulated both a compatibilist and two libertarian theories of free will in order to give both sides his best shot at a good theory of free will. Moreover, his compatibilist theory requires more than just self-control (see below for more).

dards or principles (Buss 2008, § 2).[24] One view of this sort is that an agent acts freely 'if, with other conditions satisfied, she acts from values and principles that she has neither acquired nor sustains through any process that bypasses her capacities for critical evaluation with respect to them' (Levy & McKenna 2009, 109), or, to formulate it positively, if she is 'capable of reflecting critically upon her desires, and of altering them in light of this reflection' (Buss 2008, § 2, note 5; see also McKenna 2009b, § F). [25] [26] Three other such accounts emphasize the importance of, respectively, the ability to critically evaluate norms and to make one's decisions by way of reasoning[27], various introspective, imaginative, reasoning and volitional skills[28] and the capacity to modify one's preferences in light of one's 'uncompelled reflection'[29] (Buss 2008, § 2, note 5).[30] Determinism does not imply that one cannot infer that certain things follow from one's beliefs and desires or that one cannot be moved to action by that fact, so responsiveness to reasoning is compatible with it.

Finally, there are theories that explain free will in terms of an agent's ***responsiveness to reasons***, i.e., to appropriate rational considerations, including moral ones. On this view, an action is free if it issues 'from volitional features of agency that are sensitive to an appropriate range of reasons' (McKenna 2009a, § 5.5) for and against acting in a certain way. As I said in note 23, this account of free will plausibly presupposes responsiveness to reasoning. But it goes beyond that in virtue of specifying that an agent has to be receptive for and reactive to *certain kinds* of reasons – namely, rational (including moral) ones (Buss 2008, § 2; McKenna 2009a, § 5.5; McKenna 2011, 175, 177-178, 187). There are at least seven theories that have a place within this general framework. First, it has been argued that the freedom of an action requires that it is *brought about by psychological processes that are at least moderately responsive to reasons*. Define a "mechanism" as the 'psychological processes' (McKenna 2011, 189) or 'range of agential characteristics' (McKenna 2009a, § 5.5.3) that bring about an action; and assume that a mechanism is *strongly* responsive to reasons if it will cause the agent to act otherwise when she is presented with sufficient reason to do so; and that it is *weakly* responsive to reasons if it will cause the agent to act otherwise in the case of at least some (minimal range of) reasons to do so (McKenna 2011, 190). Now, an action is free, on this account,

[24] According to Buss (2008, § 2, note 5), '[the] importance of acquiring motives in a way that is responsive to one's own reasoning is taken for granted by most writers on autonomy'. It seems to me that the same is true for most of the writers on free will. It is certainly presupposed, for example, by accounts of free will that stress "responsiveness to reasons" (see below), and arguably it is also presupposed by the view of the ability to do otherwise as not having one's agency bypassed as well as by most (but not all) of the "mesh freedoms" (see above).

[25] According to Levy & McKenna (2009, 109), this view has been defended by Alfred Mele.

[26] To be sure, no philosopher believes that either of these conditions is *sufficient* for free will. Rather, as I said above, they merely specify *necessary* conditions for free will. So, for its proponents, free will requires more than the above.

[27] According to Buss (2008, § 2), this view has been defended by Robert Young.

[28] According to Buss (2008, § 2), this view has been defended by Diana Tietjens Meyers.

[29] According to Buss (2008, § 2), this view has been defended by Gerald Dworkin.

[30] All of these accounts are accounts not of free will but of personal autonomy. Still, as with Bratman, they can be fruitfully interpreted as such, so that is what I will do in this thesis.

if it is caused by a mechanism that is *moderately* responsive to 'a rich pattern of reasons (…) that hang together rationally (…) and fit a coherent or sane pattern' (McKenna 2009b, § B.1) And this means that it is regularly *receptive* to reasons (i.e., the agent regularly comes up with, recognizes and evaluates the reasons) and at least weakly *reactive* to reasons (i.e., there is some possible world, at least, where the mechanism causes the agent to do otherwise when she is presented with a sufficient reason to do so, which 'shows that the same kind of mechanism *can* react differently to any reasons to do otherwise' (McKenna 2011, 191, italics SG; see also Levy & McKenna 2009, 113; McKenna 2009a, § 5.5.3; McKenna 2009b, § B; McKenna 2011, 189-191; O'Connor 2010, § 3.1).[31] Second, it has been defended that it is a necessary condition of an agent's being free that she is a moral agent with stable moral beliefs and deliberative principles who is capable of deliberative control. More precisely, on this view, an agent is able to act freely if (i) she has an "evaluative scheme", which includes normative (including moral) standards, desires, beliefs and plans that express the agent's long-term ends or goals, deliberative principles and the motivation to act in accordance with these components, so that the agent is minimally morally competent (Haji 2009, 149-150), (ii) she has the capacity to apply her normative standards within the process of weighing reasons (150), (iii) she is able to act on at least some of her intentions, decisions and choices (ibid.), and (iv) she comes 'to acquire an authentic [initial evaluative] scheme [i.e., the scheme that a child initially acquires] by first being instilled by others in her community with morally stable beliefs and rational principles of critical evaluation, and then [acquires an authentic evolved evaluative scheme] by being given the chance to form her own schema arising from this one' (Levy & McKenna 2009, 109; see also Haji, 2009, 151-158)), that is, by being given the chance to modify or change one's initial scheme under one's own deliberative control, by evaluating the reasons for and against the changes (Haji, 2009, 156-157).[32] In many ways, this view is very close to being a "responsiveness to reasoning" account; but in virtue of setting the substantive requirements that an agent has *moral* standards and that her moral beliefs and deliberative principles are *stable*, it goes beyond that. A third view conceives of free will as consisting in *the ability to do the right thing for the right reasons*. More precisely, it states that free will consists in the ability to appreciate (be receptive to) and act in accordance with (be reactive to) reasons that derive from "the True and the Good" (Kane 2011, 15; McKenna 2009a, § 5.4; McKenna 2011, 187).[33] A fourth account states that 'free will consists in *the ability of a person to control her conduct on the basis of rational considerations* through means that arise from, or are subject to, critical self-evaluation, self-adjusting and self-monitoring' (McKenna 2009a, § 5.2.3, italics SG). In other words, it consists in (i) coming up with good reasons and evaluating one's beliefs, desires and con-

[31] According to McKenna (2011, 189-191), this view has been defended by John Martin Fischer and Mark Ravizza. For them, though, free will requires more than responsiveness to reasons alone (see note 16 below).

[32] This view has been defended by Ishtiyaque Haji (2009, 148-158).

[33] According to McKenna (2011, 187), this view has been defended by Susan Wolf.

duct in light of these ideas (receptivity), (ii) rational self-adjustment in light of these reasons (reactivity), and (iii) self-monitoring in order to ensure that one acts in accordance with the reasons (Kane 2011, 16; McKenna 2009a, § 5.2.3).[34] A fifth theory understands free will as *being responsive to standards of right belief and desire*. More precisely, it states that free will consists in *recognizing* standards or norms of right belief and desire (which includes being open-minded with respect to (being prepared to be convinced by) the good evidence that other people might offer), and in appropriately *adjusting* one's beliefs and desires in light of these standards (McKenna 2009b, § G.3; McKenna 2011, 196).[35] Finally, a sixth account 'characterizes free will in terms of *the ability to make choices on the basis of reasons*' (McKenna 2011, 196, italics SG)[36]; and a seventh view understands free will as *normative competence*, that is, it understands free will 'in terms of (1) the power to grasp and apply moral reasons, and (2) the power to control or regulate behavior in light of such reasons' (Russell 2011, 210, italics SG)[37]. Since determinism is neutral with respect to whether agents can be responsive to rational considerations, responsiveness to reasons is compatible with it.

2.3 Theories of Moral Responsibility

Before proceeding to give an outline of the various theories of moral responsibility I shall very briefly say something about moral responsibility in general. According to Fischer (2010, 310), being morally responsible roughly consists, at the most abstract level, in 'being *accessible* to or *an appropriate target for* certain distinctively normative responses'. Some philosophers have argued that the specification of these normative responses yields different *theories* or *conceptions* of moral responsibility[38]; but others have argued that it merely designates different *aspects* of the more general and abstract concept of moral responsibility (Clarke 2010, 265; Eshleman 2009, § 2.2). It seems to me, however, that neither of them is right: all that could determine what “moral responsibility” means in this context is precisely their shared use and understanding of the concept; so, given that agreement

[34] According to McKenna (2009a, § 5.2.3), this view has been defended by Daniel Dennett. He, though (as McKenna also explains), is somewhat, well, atypical for a compatibilist, in that he denies that our thoughts have intrinsic intentionality (i.e., that they are intrinsically “about” something). Rather, he argues, adopting “the intentional stance” towards a system (i.e., a stance that employs folk psychological notions, such as beliefs, intentions and desires, in contrast, in particular, to a stance that strives to explain all systems mechanistically, in terms of the processes that cause the systems to move from one physical state into another) is proper insofar as it has utility in understanding, predicting and interacting with the system. So, presumably, he would deny that our actions “really” are caused by our deliberations (at least as we normally understand them) about what we should do. Rather, seeing ourselves as rational deliberators simply is the logical consequence of our adopting the intentional stance towards ourselves. Similarly, humans are not “really” morally responsible agents. Rather, they can properly be *regarded* as morally responsible agents if interpreting them in that way (adopting the “personal stance” towards them) has utility.

[35] According to McKenna (2009b, § G.3), this view has been defended by Philip Pettit and Michael Smith.

[36] According to McKenna (2011, 196), this view has been defended by Kadri Vihvelin.

[37] According to Russell (2011, 210), this view has been defended by R.J. Wallace as a Kantian theory of “agency” or “normative competence”. It can easily be interpreted as a reasons-responsive theory of free will, however.

[38] According to Clarke (2010, 264), this view has been defended by Tim Scanlon, Gary Watson and Michael Zimmerman.

on what it means is absent, there simply is no fact of the matter as to what it means. Even so, in what follows I shall, for convenience, *interpret* the specification of the proposed normative responses as the specification of distinguishable (but related) *kinds* of moral responsibility.

First, moral responsibility has been thought by some to consists in ***attributability***[39]. On this view, which simply is the mirror view with respect to moral responsibility of the last kind of "mesh freedom" that I have discussed, 'you are morally responsible for something just in case it is attributable to you as a basis for moral assessment of you' (Clarke 2010, 264), that is, just in case it reveals or expresses your moral nature or character or attitudes and values. That is, being morally responsible for some item consists in that item being expressive of the agent 'in a way that would make the agent a sensible or appropriate target for distinctively moral judgments' (Fischer 2010, 311) or 'for ascriptions of certain ethical predicates' (Fischer 2007, 333), in particular judgments with respect to which *character traits* or *vices and virtues* of an agent some item expresses and that measure the character trait(s) that some item expresses against some moral standard. As with the last "mesh freedom", this means that it is possible to be morally responsible for something that is not *voluntary* (not within one's control, in the sense that one cannot influence its occurrence, at least in the short term; e.g., having certain thoughts or feelings or failing to remember certain things). All that being responsible for something requires is that it expresses one's *judgment-sensitive* attitudes, which are attitudes that are sensitive to judgment in *ideally rational* agents (Clarke 2010, 264; Eshleman 2010, § 2.2; Fischer 2007, 333; Fischer 2010, 311; Levy & McKenna 2009, 115; McKenna 2009b, § G.1).[40] Determinism does not imply that one's actions can never be expressive of one's judgment-sensitive attitudes in this way, so this kind of moral responsibility is compatible with it.

Second, moral responsibility has been understood by some as ***appraisability***. These philosophers argue that being morally responsible consists in being the sort of person whose 'moral standing or record as a person is affected [i.e., enhanced or diminished] by some episode in, or aspect of, one's life' (Haji 2009, 17). In metaphorical terms, being morally responsible consists in having a "moral ledger" on which (positive, negative or neutral) marks can be made: being responsible for something good leads to a 'credit or luster' (Clarke 2010, 264) being assigned to one's moral ledger (which enhances one's moral standing); and being responsible for something bad leads to a 'debit or blemish' (ibid.) being added to one's ledger (which diminishes one's moral standing). This kind of moral responsibility evidently does presuppose that the positive, negative or neutral mark that is entered into one's moral ledger can be properly attributed to oneself. But it differs from the "attributability" kind

[39] All terms for the different kinds of moral responsibility, and the classification of the sorts of moral responsibility there are that goes along with the use of those terms, come from Clarke (2010, 264-265).

[40] As I've said when discussing the relevant "mesh" freedom, this view of what moral responsibility consists in (and what it requires) has been defended by T.M. Scanlon and Angela Smith. Once again, however, note that they argue that *holding* agents responsible (blaming and praising them) *does* require a certain degree of voluntariness or control.

of responsibility 'in holding that, of the variety of moral assessments of agents, only a narrow range are ascriptions of moral responsibility' (Clarke 2010, 264) – for instance, ones that express what the agent identifies with or values, or ones that express her larger plans (Clarke 2010, 264; Eshleman 2010, § 2.2; Fischer 2007, 333; Haji 2009, 17)[41]. Is this kind of moral responsibility compatible with determinism? Well, it seems that to me that much depends on how strong an interpretation the concepts of "moral standing" and "having a moral ledger" are given. If they are only meant to denote that a more narrow range of items can properly be attributed to an agent's moral character or self than the first conception of moral responsibility says, then it seems plausible to me that this kind of moral responsibility is compatible with determinism. But if, as Haji (2009, 17) says[42], having one's moral standing enhanced by something renders one *worthy* (or *deserving*) of praise and having one's moral standing diminished by something makes one *worthy* of blame (i.e., if having a moral standing or "ledger" makes it the case that one can be *worthy* or *deserving* of praise and blame), then it is arguable that moral responsibility of this sort is *not* compatible with one's being causally determined by factors that are beyond one's control. (I will expand on this below.) In my view, however, that simply isn't true. After all, if it is merely *coherent*, as seems intuitively evident to me, to claim that a person can have a moral standing without it being possible that she is worthy of praise and blame, then the argument that the former logically *implies* the latter has failed.

A third view of moral responsibility states that it consists in ***answerability***, i.e., that being morally responsible for something consists in being *answerable* for that thing (Clarke 2010, 264), in the sense that it would 'not be inappropriate' (Fischer 2007, 333) or 'legitimate' (Pereboom 2011, 408) to expect the agent, if possible, to *explain* or *justify* (to answer critical question about), say, her choice or action, and, if that is not possible, to expect her to acknowledge that she has done something wrong (Clarke 2010, 264; Fischer 2007, 333; Pereboom 2011, 408).[43][44] As with moral responsibility in the sense of *appraisability*, being answerable for something presupposes that it can properly be attributed to oneself (presumably in a relatively narrow sense, as it seems counterintuitive to be asked to explain or justify something that is not voluntary); but since this sort of moral responsibility denies that it is *constitutive* of being morally responsible for something that it can properly be attributed to oneself, the former cannot be reduced to the latter. Similarly, it could be argued (as Fischer (2007, 333) does)

[41] Clarke (2010, 264), Fischer (2007, 333) and Haji (2009, 17) all refer to (or even cite) Michael Zimmerman, so presumably he is the most prominent defender of this view (although none of them explicitly says that he is).

[42] To be fair to Haji, however, he merely *describes* this view of moral responsibility, without endorsing it.

[43] According to Clarke (2010, 264), T.M. Scanlon has argued that attributability *includes* answerability, and, therefore, that answerability is an *aspect* of moral responsibility. Pereboom (2011, 408) also refers to Hilary Bok when discussing this view; and Fischer (2007, 333) refers to Marina Oshana.

[44] It seems to me that this view of moral responsibility fits rather well with "responsiveness to reasoning" and especially "responsiveness to reasons" accounts of free will, as those freedoms allow one to explain or justify one's behavior. None of the writers that I have read make this connection, however, so don't take my word for it.

that being answerable for one's choices and actions, say, *implies* that one has a moral ledger; but even if this so (counter to my intuitions), being answerable for something still is distinguishable from its enhancing or diminishing one's moral status, and that is all that the account needs. At a first glance, it seems rather self-evident to me that “being answerable for something” is compatible with causal determinism, for determinism seems to be entirely neutral with respect to whether one can (or cannot) be reasonably be expected to explain or justify, say, one's behavior. *If*, however, Fischer is right that being morally responsible in this sense implies having a moral ledger, and *if* having moral standing is a sufficient condition of being worthy of praise and blame for one's actions, *then* it is arguable that this kind of moral responsibility is incompatible with there being sufficient conditions for one's choices and actions in the distant past. But those “ifs” don't strike me as plausible: as it seems possible to think of a possible world wherein agents have moral standing but cannot be worthy of praise and blame, so it seems coherent to imagine a world in which it is appropriate to expect agents to explain or justify their behavior without their having a moral ledger with marks on it (e.g., a world – perhaps ours – in which the required personal identity over time does not obtain).

Finally, it has been argued that moral responsibility consists in ***accountability***. Being morally responsible for an item in this sense consists in it being appropriate that one is *held to account* for that item. In other words, it consists in being an appropriate or apt target (object, candidate) for the so-called “reactive attitudes”, which are evaluative attitudes involving certain emotional reactions (Haji 2009, 16; McKenna 2009, § 4.3.1), such as admiration, forgiveness, guilt, hatred, love, moral sadness, respect, indignation and resentment, to people's behavior, in particular behavior that expresses the good or ill will or indifference (Haji 2009, 16) of others towards ourselves or others, or of ourselves toward others. When they are indeed reactions to such behavior, they express the demand that other people (and we, too) display a 'reasonable degree' of good will towards us or towards other people (16-17). So, in a nutshell, 'to be [morally] responsible is to be part of a moral community in which people can appropriately adopt the reactive attitudes toward one another' (17; see also Clarke 2010, 264-265; Eshleman 2009, § 2.2; Fischer 2007, 333; Fischer 2010, 311; Kane 2011, 15; and McKenna 2009, § 4.3, § 5.6). At this point, however, things are becoming complex. To begin with, two broad interpretations of what “appropriateness” consists in must be distinguished: a social constructivist interpretation and a moral interpretation. On the first interpretation, 'those in the moral community determine the conditions for when a person is or is not a morally responsible agent, as well as whether a person is or is not responsible for some bit of conduct' (McKenna 2009a, § 5.6). So whether a person is an appropriate object of the reactive attitudes is not determined by the underlying nature of the person. Rather, this simply 'derives from the normative considerations embraced by the members of the community holding people responsible' (ibid.), i.e., the conditions under which it is appropriate to direct one's reactive attitudes at an individual are *socially constructed* by the members of the moral community. Determinism obviously does not threaten this kind of moral responsibility, because moral responsibility can simply be constructed in such a way that it is compatible with determinism. The second interpretation states that whether an individual is an appropriate candidate for the reactive attitudes is determined by *moral principles* about when a person is (or is not) an appropriate target for the reactive attitudes. At most, moral norms concerning when it is appropriate to direct one's reactive attitudes at a person are *reflected* in the practices of a moral community. But a moral community *can* 'respond to a group of persons inappropriately' (ibid.). This kind of interpretation can be worked out in (at least) three ways. First, it could be argued that an agent is an appropriate target for some reactive attitude if and only if she is *worthy* of that reaction, in the sense that she *deserves* or *merits* it for non-consequentialist reasons (e.g., because she voluntarily has some bad state of mind, such as a bad will or bad intentions, or because she voluntarily aims at an end that is morally wrong in some way), given her behavior and/or character traits (Eshleman 2009, § 1, § 2). In my view, this kind of moral responsibility is *not* compatible with one's behavior and character being determined by factors that are beyond one's control; and neither is it compatible with one's behavior and mental nature being probabilistically caused by factors that one does not control, or with those things not being caused at all. So, at first sight, it seems to be impossible (or close enough). I will expand on this below. Second, it can be argued that the general practice of holding each other accountable for each other's behavior is legitimate; but that the appropriateness of directing a certain reactive attitude at a certain individual in certain circumstances should be guided by considerations of fairness (that are independent of the above-mentioned notion of desert). For instance, it could be argued that holding each other accountable for each other's behavior is in general morally legitimate in virtue of the role it plays, by informing each other about and enforcing certain desirable moral norms, in encouraging the development and expression of virtuous character traits; but that this practice has to be constrained by considerations of fairness, in the sense that in some circumstances, and for some kinds of individuals, it is *not* appropriate to direct one's reactive attitudes at the relevant person (my outline of this account is roughly inspired by the Aristotelian account of moral responsibility that Williams (2006, § 3.b) describes). This could be so, e.g., because that individual did not *want* to perform the action (but was forced to perform it by something that is, in some sense, alien or external to her), because she did not *know* what she was doing or what she was bringing about, or because she is mentally ill or morally underdeveloped in some way. On this view, in short, our responsibility practices are taken as a legitimate given with regard to regular persons and situations in virtue of the essential role they play in our lives, but questions of the appropriateness of reactive attitudes still arise for individuals and situations that are abnormal in some way. Determinism does not entail that it is impossible to encourage the development of virtues (and to discourage the development of vices), and neither does it rule out considerations of fairness, so this kind of moral responsibility is compatible with it. Finally, it could be argued that both the rightness of the practice of holding each other accountable for each other's behavior in general *and* the rightness of directing a certain reactive attitude at a certain individual in a certain situation are determined by whether they have (sufficiently) good (expectable) consequences).[45] The general practice is legitimate, on this account, in virtue of the fact that it reliably produces (sufficiently) good (expectable) consequences. It is not difficult to see why: humans generally are very sensitive to the reactive attitudes of others (and their more retributive counterparts: praise, credit and reward; and blame, sanctions and punishment). And that means that which choices they make, which actions they perform and which character traits they develop can be influenced, guided and reinforced in desirable ways through (holding out the prospect of) directing certain reactive attitudes (etc.) at individuals for making certain choices, behaving in certain ways and possessing certain kinds of character traits. Note that this interpretation, unlike the previous one, mentions more kinds of consequences than the development of virtuous character traits alone. For instance, it may reasonably be argued that some (or even many) individuals simply are hopeless with respect to the exercise and development of their deliberative capacities in order to perfect their responsiveness to reasons: they may have the wrong kind of genes, or may have had the wrong kind of childhood, or they may live in the wrong kinds of circumstances – in any case, they don't have a chance of developing virtuous character traits. Still, their *behavior* can in virtually all cases be influenced in important ways: they may be deterred from committing crimes by holding out the prospect of resentment and indignation and blame, sanctions and punishment; and they may be encouraged to perform certain actions by holding out the prospect of admiration, love and respect and praise, rewards and credit. These individuals may perform those actions for all the wrong reason, and not with the consistency and thoroughness of a virtuous agent. But they do perform them, and that's better than nothing. As if all this isn't enough, there are two more arguments in favor of the general practice, proposed (as arguments for compatibilism) by the famous (nonconsequentialist) philosopher Peter Strawson (for these arguments, see Deigh 2011, 204-205; Eshleman 2009, § 2.1; McKenna 2009, § 4.3; Russell 2011, 202-203; and Timpe 2006, § 5.c). First, he argued that, for us humans, it simply is not psychologically possible to stop having reactive attitudes towards the behavior of others: having those attitudes is a necessary consequence of engaging in interpersonal relationships, and we simply cannot do without those. In my view, Strawson is absolutely right about this: humans simply cannot entirely suspend or abandon their reactive attitudes insofar as they want to get into interpersonal relations

[45] In what follows, I will provide an outline of one such account. Many others could be given, however.

with other individuals. It *is*, of course, possible (in principle at least) to abandon or at least mitigate *some* reactive attitudes (e.g., hatred or resentment), but suspending all of them is not. And, second, Strawson argued that even if we *were* capable of giving up the reactive attitudes, doing so would diminish the richness and quality of human relationships, and of human life in general, to such an extent that it would be blatantly irrational to do so. On this point, too, I firmly agree with Strawson. Although the argument certainly does not succeed in establishing that we should not give up *any* of the reactive attitudes, it plausibly does show that giving up *all* of them would cause great losses to human life; and for a consequentialist, that implies that abandoning them is wrong.

On to the second element. I shall first consider which individuals are, on this view, appropriate objects of the reactive attitudes (etc.). According to Deigh (2011, 195, 207-210), the classical utilitarians (in particular Jeremy Bentham) reasoned as follows – for punishment, but it can easily be extended to the reactive attitudes. They first note that being punished reliably produces negative experiences – and that is a bad thing, for those are disvaluable. Still, holding out the prospect of punishment deters some individuals from performing actions that cause negative experiences; and preventing those actions probably prevents more negative experiences than (the prospect of) punishment causes; so punishment is, all things considered, a good thing (208). Some individuals, however, simply are not (or barely) capable of modifying their behavior in response to sanctions and rewards (e.g., because they are mentally ill, or addicted to drugs), so that they cannot 'reliably determine which, among the different actions open to them in the circumstances they face, is the best one to do (…) [or which one] they have most reason to do and then to act accordingly' (195). For these individuals, the threat of punishment simply is not effective: it does not, and cannot, prevent their actions. For that reason, they should not be punished: that would only serve to cause negative experiences for those individuals without producing any positive ones. So, in a nutshell, it is only morally right to punish an individual if she is capable of rationally determining which of several actions is in her best interest: if she cannot do so, then punishing her is useless. This certainly seems *prima facie* plausible to me: why punish an individual if doing so does not at all deter her (or others like her)? Much better to simply isolate or (if possible) treat and rehabilitate that individual. And the same point applies to the (prospect of the) reactive attitudes: if an individual is not (or barely) sensitive to the prospect of the) reactive attitudes (because she is insufficiently developed or mentally ill), then it would seem to be superfluous – indeed, barbarous – to direct one's negative reactive attitudes at her. Instead, she should be treated or educated – or, in the worst case scenario, quarantined.

Having considered what kinds of *individuals* are appropriate objects of the reactive attitudes (etc.), I shall now consider in what kinds of *circumstances* the reactive attitudes are appropriate. In my view, it is quite plausible that it is wrong to direct a negative reactive attitude at an individual for performing an action, say, if she did not perform it voluntarily (if she was forced to perform it; if her action was an accident or unintentional mistake; if she had no, or very little, control over whether to perform it or not; and so on) and/or knowingly (if she was not aware of the nature or significance of her action and its consequences). After all, in those cases the individual did not really want to do what she did (she did it against her will); and given that fact, she could not have been deterred by (the prospect of) negative reactive attitudes (etc.). And the same might be true for circumstances in which an individual cannot reasonably be expected not to perform an action, such as when not performing that action would cause great harm to that individual (Deigh 2011, 209). It has been objected[46], however, that this is not true (209-210): directing one's reactive attitudes (etc.) at an individual in those circumstances may, after all, at least for some sorts of actions (e.g., car accidents), encourage agents to become more cautious and careful than they would otherwise have been, and it may deter some agents from performing certain actions hoping that they can get away with it by pleading ignorance, mistake, and so on. In response, I have three answers. First, given that I have argued that negative experiences have more disvalue than positive experiences have value, it is at least *prima facie* plausible that, in many cases at least, it will have better consequences to acknowledge the above mentioned excusing conditions as valid. Second, moreover, note that many people *as a matter of fact* regard these conditions as providing (once again, at least in many cases) valid excuses. Given that fact, it seems to be, at least in our current circumstances, recommendable to (often) treat them as such as well – not only because going counter to people's moral expectations hurts them, but also because directing your negative reactive attitudes (etc.) at those individuals in those circumstances might well not be effective (it might only serve to get them angry and frustrated) and because it might lead them, if you do it often, to opt out of your relation (whatever relation it is). Finally, it may be thought that, given that my kind of consequentialism emphasizes expectability, and given that there are, as yet, no knock down arguments either way (which means that it is not, at this point, foreseeable which of the several options available to us is the sufficiently good one), it is best to stick with the status quo and do what it is common to do, because the consequences of our current practices are at least known (and not awful). So it seems to me that we have, at least until convincing contradictory evidence appears, good reason to roughly respect these excusing practices.

Finally, I shall consider which reactive attitudes it is right to direct at individuals, and in how far. In my view, a distinction is relevant here between the “positive” (admiration, forgiveness, respect, and perhaps praise, rewards and credit) and the “negative” (guilt, hatred, moral sadness, indignation, resentment, and perhaps blame, sanctions and punishment) reactive attitudes. It seems to me that the value and widespread appropriateness of the positive reactive attitudes is beyond dispute. Sure, they

[46] According to Deigh (2011, 209-210), this argument has been given (with respect to punishment) by H.L.A. Hart.

have to be directed at the right individuals in the right circumstances, but if they are, then they are no doubt legitimate: they produce positive experiences in those individuals (it feels good to be admired, forgiven, loved, respected, praised, rewarded or given credit); and, if done right, they steer them in the right direction and encourage the development of certain valuable character traits. The adoption of the negative reactive attitudes is more difficult to justify, however. To see why, note first that I believe (feel, have the inuition, or whatever) that negative experiences are more disvaluable than positive ones are valuable. That means that we should at least make *more* use of the positive reactive attitudes than the negative ones for modifying others' behavior, in order to avoid producing disvalue (provided, of course, as I will assume here, that positive reactive attitudes are at least as effective as negative reactive attitudes). Furthermore, some of the negative reactive attitudes – hatred, indignation and resentment in particular – seem rather vicious: adopting them seems hard to combine with making a realistic assessment of a situation, and may well encourage illegitimate quarrels, political polarization and political deadlock, and even wars – in short, they often encourage individuals to perform actions that are irrational, immoral, unreasonable, ineffective, harmful and plainly indefensible, and that definitely reduces their value. Moreover, certain valuable character traits (i.e., traits that either are valuable in themselves or because they have good consequences), such as a tendency to emphatize with and to be kind and beneficent to others, probably make it very difficult to feel these emotions. Finally, diminishing the hatred, indignation and resentment that one feels probably also has good consequences for one's own life: since all of these attitudes constitute negative experiences, cutting down on them plausibly reduces the amount of negative experiences that you have; and certain valuable character traits, such as the tendency to accept that individuals have a certain character and to strive to prudently deal with that fact (rather than trying to change them, often to no avail), may well bring it about that one feels less hatred, indignation and resentment. It is true that these emotions can also inspire individuals to do good things (e.g., to bring down a corrupt and repressive government, or to vigorously work in order to end some injustice), but these can probably be achieved without them as well – and without their harmful by-effects. So it seems plausible to me that at least a significant reduction in the presence of hatred, indignation and resentment would be a good thing. What about guilt and moral sadness? Well, given that they play an important role in moving individuals to strive to improve their character and conduct, and in the latter case the behaviors of others too, given that they are not even close to as prevalent as indignation, resentment and hatred, and given that they do not seem to have large and vicious by-effects, their presence probably is not in need of significant reduction. In light of the disvalue attached negative experiences, one should of course take care not to have *too* much of them; but for most individuals at least, that probably will not be much of an issue. What about blame, sanctions and punishment? To begin with, note that these are all essential to our lives – all of them are needed in order to discourage certain individuals from performing certain actions that have bad consequences. Even so, it does seem to be the case that our blaming practices are often somewhat excessive: when something has gone wrong, people often immediately start fingerpointing – even if they don't have adequate information, and even if doing so is not reasonable, prudent or effective. So it probably would be a good thing to temper down our blaming practices a bit. And our practices with regard to sanctions and punishment? Well, I don't know. Are there too many sanctions or too few? Are people punished too often, or too little? I simply don't know. I *do* know that our punishment practices are rather inefficient with respect to rehabilitation, and that many of our sanctions are not very effective in deterring very influential individuals from performing bad actions in order to get at more profit or power; and I believe that the design of sanctions and punishment should be directed by the amount of value that they produce; but that, of course, is not much. So, for now at least, I will remain neutral on precisely what shape our punishment practices and sanctions should have.

Before closing this section, there is one last distinction I need to make. Many philosophers believe that being an appropriate object of the reactive attitudes *implies* or even *includes* being an appropriate object of praise (credit and reward) and blame (sanctions and punishment), that is, they believe that some of the reactive attitudes are inherently retributive. But, according to Clarke (2010, 365), some philosophers 'deny that these attitudes have a retributive element'.[47] Keep that in mind.

[47] Clarke refers to T.M. Scanlon as one of those philosophers.

Chapter 3: In Defence of Evaluative Compatibilism

Now that I have provided an overview of the most prominent incompatibilist and compatibilist kinds of freedom that have been proposed as accounts of free will and of the most prominent kinds of responsibility that have been proposed as accounts of moral responsibility, I can proceed to evaluate them in order to establish evaluative compatibilism. Before being able to do that, however, I need to make one rather important note – that I will limit my attention to considering whether all these kinds of freedom and responsibility are valuable *from a rational point of view*. For instance, it could be the case that some people (or many people) value an "absolutist" kind of freedom, and that it would cause them pain if they discovered that there is no such thing, even though they *have no good reason* for valuing that kind of freedom. Given that I accord intrinsic value to the avoidance of negative experiences, and given that the nonexistence of this kind of freedom would (if known) bring it about that some (or many) individuals have negative experiences, it could be argued that my theory of value implies that this kind of freedom is at least instrumentally valuable. Even so, I will ignore this complication, for two reasons. First, even if the known nonexistence of some otherwise uninteresting kind of freedom or responsibility would indeed bring it about that some (or many) people will have negative experiences, this will probably be a short-term process: after some period of time, they will probably conclude that is not that big a deal that the relevant kind of freedom or responsibility does not exist. Some (or many) of them might come to this conclusion on the basis of rational deliberation, but that is not even necessary: given that humans are biased towards optimism, they could equally well come to this conclusion on the basis of a process of *rationalization*, i.e., on the basis of a process in which they (perhaps unconsciously) *search for* reasons that serve to reconcile them with their predicament. In either case, it will only be a matter of time until the negative experiences are over – a period, moreover, that might well encourage the development of the agent's responsiveness to reasons. Of course, the avoidance of some negative experiences still has some value, but not enough (in the long run) to give it much thought. Consider, by comparison, a young theist who discovers that it is highly improbable that there is an omnipotent, omniscient and omnibenevolent being (God). At first, this no doubt will be a painful realization. But it will probably not bother him *that* long. And even if it does bother him for quite some time, will it bother his children? Not likely. I have been raised by atheist parents, and I have *not ever* worried about the existence of God – I simply *don't care*. Sure, it would be fantastic if God would exist and I would happily live in heaven for eternity, but things can still be quite fantastic without Him. Many people have come to that conclusion. And if they can complacently recognize His nonexistence, then they should certainly be able to reconcile themselves with the nonexistence of something of very little importance. Note, however, that there is one important disanalogy between the “God”-case and the “free will”-case: whereas people genuinely *care* about the existence of God – they devote a lot of time and effort towards honoring him, e.g., by preying, or going to church, or studying whatever book they consider sacred – it is not so clear that this is so for any of the incompatibilist (or otherwise “exotic”) freedoms and responsibilities. It seems plausible to me that most “ordinary” people (and many philosophers not specialized in free will and moral responsibility) simply do not give free will and moral responsibility much thought – they may read an article about those topics in a newspaper and may occasionally converse about them with a friend, but that probably is it. Moreover, it may matter a lot how the news that a certain kind of freedom or responsibility does not exist is framed. Even if nonphilosophers *do* care more than I consider probable about the existence of some incompatibilist freedom or responsibility, one might wonder whether they would really lose much sleep over their nonexistence if one would not emphasize the kind of freedom or responsibility that they *don't* have (“you don't have free will!”), but if one would instead insist that there are many valuable freedoms and responsibilities that they *do* have (“you may not have freedom *a*, but you *do* have the freedoms *b*, *c*, *d*, and *e*, and look how wonderful they are!”). So, in short, it seems plausible to me that knowing about the nonexistence of some incompatibilist freedom or responsibility would not do much harm in the long run. Second, there is an *epistemic* reason for not directing one's attention at whether some freedoms, at least, are valuable because they are valued – namely, because at this point *we have no idea* which kinds of freedom are considered valuable by nonphilosophers. It could even be argued that *they* don't know this either, in the sense that they simply haven't thought about it. To be fair, this argument does not work for moral responsibility: in that case, it *does* seem quite obvious that many nonphilosophers care about the existence of one kind of moral responsibility that I believe to be incompatible with determinism – the desert-entailing “accountability” one. For moral responsibility, therefore, I will have to rely on the first argument. In short, (i) the knowledge that some incompatibilist freedom or responsibility does not exist would probably not bring it about that many individuals will enduringly suffer from negative experiences, so those experiences are no reason to ascribe much value to the relevant freedom, and (ii) we simply do not know which kinds of freedom nonphilosophers value, so ascribing value to some freedom on that basis would, in my view, be premature and unduly speculative. Together, these arguments constitute good reasons, I believe, for not including this complication in my evaluation of the various freedoms and responsibilities.[48]

[48] According to Kane (2011, 26), Saul Smilansky defends a third strategy: if knowing that some freedom does not exist (as opposed to the nonexistence of that freedom itself), would have dire consequences, he argues, then this means that we should simply *hide* the nonexistence of that freedom from the individuals that would be negatively affected by knowing about it – not in the sense that we have to *induce* the belief that they have that kind of freedom in them, but in the sense that we should not *interfere* with their belief that they possess that sort of freedom but leave it intact. In my view, however, this strategy is not sustainable: if science will go on progressing as it does, and if information will continue to spread faster and faster, then it will probably prove to be impossible to prevent that information

Having noted that important complication, I can now go on to evaluate the variety of freedoms and responsibilities that I have described above based on the theory of value and the normative ethical theory that I endorse, in order to establish that evaluative compatibilism is correct. In section 3.1, I will argue that none of the incompatibilist kinds of freedom that have been proposed as (parts of) accounts of free will is valuable. In section 3.2, I will argue that some of the kinds of freedom that have been proposed as (parts of) accounts of free will and are compatible with determinism are valuable. In section 3.3, I shall argue that none of the kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility and that are incompatible with determinism is valuable. Finally, I will argue, in section 3.4, that some of the compatibilist kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility are valuable.

3.1 Evaluating the Incompatibilist Freedoms

As I said above, I know of the existence of three prominent incompatibilist kinds of freedom that have been proposed as (parts of) accounts of free will. First, there are ***absolutist*** kinds of freedom; and there at least two freedoms of this sort.[49] The first of these consists in intentionally bringing it about, from scratch, that one is the way one is in certain crucial mental respects (i.e., that one has a certain mental nature, or a certain character). Based on the theory of value that I endorse, it is hard to see what value this kind of freedom could have: it does not consist in a positive experience; there is no reason to suppose that it will lead to the avoidance of negative experiences; and neither does having this kind of freedom consist in or contribute to the perfection of one's responsiveness to reasons. On my theory, what matters is not that one's mental nature has a certain *origin* (that one has intentionally brought it about), but that it has good *consequences*; and there simply is no guarantee that intentionally bringing one's character into existence will have good consequences. So it seems plausible to me that this kind of freedom is not valuable. The second of these freedoms consists in bringing it about that one makes a certain choice or performs a certain action without being causally influenced by any external causal factors, including one's psychological states and character. Once again, this freedom seems to be devoid of value, because it does not seem to consist in or lead to any of the things that I have judged to be intrinsically valuable. One can, after all, easily conceive of an individual that is not causally influenced by any external causal factors, but that still consistently makes the wrong choices. Moreover, if one's psychological states are, or one's character is, good, then it may well be desirable that one is influenced (or even determined) by those factors.

from reaching the public, and to prevent that public from being convinced by the increasingly plausible results.

[49] For the sake of the argument, I will assume that both of them are coherent (counter to my intuitions).

Second, it has been argued that it is a necessary condition of the freedom of actions that they are events that are ***uncaused***. Given that the fact that an action is uncaused does not imply that it has a certain nature that is valuable, or that it has certain valuable consequences, there is no reason to ascribe any value to this kind of freedom on my theory of value. An uncaused action *could* certainly be valuable, but if it is, then it is so for reasons that are entirely independent of its being uncaused.

Finally, there is ***the libertarian ability to do otherwise***. One is able to do otherwise in this sense if one is able to (if it is possible for one) to do otherwise (to perform another action than the one that one actually performs) given *exactly* the same past and *exactly* the same laws of nature. Apart from the suggestion that the libertarian ability to do otherwise is required for moral responsibility in the desert-entailing "accountability" sense, I have encountered two sorts of explicit suggestions as to why this ability is valuable; but on my theory of value, neither of them is appealing.

First, it has been suggested (e.g., in Clarke 2008, § 2.1, § 2.5) that the libertarian ability to do otherwise is valuable because its possession allows us to *make a difference*, through acting, to the course of the world. As Clarke (2008, § 2.1) puts it, if we are able to do otherwise in the libertarian sense, then 'there are things that happen that might not have, and there are things that do not happen that might have, and sometimes in performing free actions, we make a difference to which such things happen'. If you endorse my theory of value, however, then there is no reason to suppose that "making a difference" in this sense is valuable, because there is no reason to believe that bringing it about that certain things happen that might not have happened (or the other way around) will by itself ensure that *the right things* will be made to happen – and *that* is what matters. A somewhat similar suggestion is that 'if it is rational (or at least not irrational) to prefer that one's actions do not have sufficient conditions which predate their proximal deliberative springs, then it might be rational (or at least not irrational) to prefer to be a libertarian agent' (Levy & McKenna 2009, 122). But, with my values, it *is* irrational to *prefer* that there are no sufficient conditions of one's actions prior to their taking place, because on my theory the value of an action has to be assessed in terms of its intrinsic nature and in terms of its consequences rather than in terms of how it came into being.

And, second, there have been offered a variety of arguments that aim to show that the truth of determinism would somehow be in tension with deliberation. First, it has been argued 'that it is a conceptual truth that I cannot engage in deliberation if I do not believe that I have free will, in the sense that involves alternative possibilities' (Fischer 2007, 324).[50] But, as Fischer (2007, 325) says, it is hard to see why this should be so: even in a deterministic world, after all, it is likely that it will have better consequences (in certain situations) to deliberate about what we should do, because that is the only way we can figure out what it is rational or reasonable or right to do, given our values. Second, it

[50] According to Fischer (2007, 324), this argument has been defended by Richard Taylor.

has been suggested that even though it is *possible* to deliberate while believing that the libertarian ability to do otherwise does not exist, in doing so one would be *contradicting* oneself, because deliberating about whether to perform one action or another presupposes a belief that it is *possible* to perform either of those action (Fischer 2007, 325).[51] In my view, however (once again in line with Fischer), this simply is not true: weighing reasons for and against performing one of several courses of action does not presuppose a belief that one can (in the libertarian sense) perform either of them, but only that one does not know which of those actions one will perform, and that one wants to find out which of them one (as Fischer likes to put it) has sufficient reason to choose. Similarly, Clarke (2008, § 2.5) suggests that, in deliberating, some individuals inevitably experience some kind of openness or “leeway”, in the sense that 'they cannot help but presume that more than one course of action is genuinely open to them'. Since it is undesirable, he argues, to be subject to illusory experiences, for these individuals, at least, it is rational to prefer indeterminism above determinism. But it seems to me that the fact that we are subject to an illusory experience does not make it rational to prefer a world in which the relevant experience is not illusory (e.g., a world that is indeterministic), so long as having that experience does not have bad consequences. For instance, I am unavoidably subject to the illusion that the ordinary objects that I perceive are rather solid, even though they are, in fact, mostly empty space. But I see no reason for deploring that fact: it does not cause me pain; and neither does it hinder the development and expression of my deliberative capacities. Finally, it has been argued that knowing that determinism is true would render deliberation *pointless*, for the following reason. If I would know that the beliefs and desires preceding my choices and actions are causally sufficient for their occurrence (that it is not causally open to me to choose any of the alternatives that I consider), 'then I could just sit back and watch the action unfold in the same way as I do when I sit back and watch the action unfold on a movie screen' (Searle 2001, 71-72), because in that case our deliberative processes simply would not make a difference: '[t]he bodily movements were going to be exactly the same regardless of how these processes occurred' (285-286).[52] This is not true, however: if determinism is true, then our deliberative processes (probably) *do* make a difference, in the sense that the occurrence of certain decisions and bodily movements is *dependent* upon the taking place of certain rational decision-making processes (they would not have occurred if those processes would not have taken place). It simply is *not* the case, as Searle says, that determinism entails that certain actions are going to occur regardless of how my deliberations proceed (Fischer (2007, 327-331). It plausibly *is* true, as Fischer repeatedly emphasizes, that meaningful deliberation requires an *epistemic* sort of openness, the openness of *not yet knowing* which of the several decisions and

[51] According to Fischer (2007, 325), this argument has been defended by Peter van Inwagen.

[52] According to Fischer (2007, 327-330), this argument has been defended by John Searle (whom he quotes at length).

actions that I am capable of making or performing (in a compatibilist sense) I will make or perform. But given the fact that the famous “uncertainty principle” of quantum mechanics, which states that certain combinations of observable properties of quantum mechanical systems cannot be measured precisely (Hodgson 2011, 59), implies that even the behavior of elementary particles involves some (at least) predictive indeterminacy, and given the fact that deterministic chaotic systems (i.e., systems that are highly sensitive to very small changes in initial conditions (Bishop 2011, 91)) can behave highly unpredictably, this is something that we probably will never have to worry about.

On the other hand, it does not seem to be the case that determinism is preferable over indeterminism either. Compatibilists have argued for quite a long time that if an agent might have acted otherwise given exactly the same beliefs, desires and reasons, then it is a mere matter of luck or chance how she acts, because there is 'nothing about the agent prior to the decision – indeed, there is nothing about the *world* prior to that time – that accounts for the difference between her making one decision and her making the other' (Clarke 2008, § 2.2). In other words, they have argued that if the occurrence of more than one choice or action is consistent with all facts about the world and the agent in it, then which of them occurs is a mere matter of chance or luck, because which of several choices you will make or which of several actions you will perform is settled not by *you*, but by a random process that has nothing to do with you whatsoever. And it seems to me that this argument succeeds[53]: if more than one choice is causally open to an agent given all her reasons, then her reasons simply cannot explain why she makes one choice *rather than another*. But it is hard to see why this matters, for two reasons. First, luck does not necessarily have to be *bad* luck, but can also be *good* luck, so there is no reason to think that the presence of luck will have worse consequences than its absence. And, second, deterministic agents are thoroughly subject to luck as well. To be precise, they are subject to at least four kinds of luck: *resultant* luck (luck with respect to the consequences of one's actions and projects); *circumstantial* luck (luck with respect to the circumstances in which one is situated); *constitutive* luck (luck with respect to who one is, or what mental nature – what character traits, potentials, capacities and dispositions – one has); and *causal* luck (luck with respect to how one is determined by antecedent circumstances)[54] (Latus 2008, § 2b; Lippert-Rasmussen (2009), § 1; Nelkin 2008, § 1). Replace “determined” with “causally influenced in accordance with probabilistic laws”, and all these kinds of luck affect indeterministic agents as well. So whatever world one finds oneself in, one's life will unavoidably be shaped by luck in an indefinite number of ways.

[53] To be sure, several (rather unconvincing) objections have been offered to this argument; but I will not go into them.

[54] It is controversial whether this category of luck should be included, because circumstantial luck (luck in what happens to one) and constitutive luck (luck in who one is) seem to cover all that causal luck could encompass (Latus 2008, § 2b.iii; Nelkin 2008, § 1). I should also mention that the categories (although not their names) have been proposed by Thomas Nagel in the context of the “moral luck” debate (according to Latus, Nelkin and Lippert-Rasmussen).

So I see no reason, given my values, either to prefer indeterministic freedoms over deterministic freedoms, or to prefer deterministic freedoms over indeterministic freedoms. For a freedom of genuine value, therefore, I will have to look elsewhere than mere determinism or indeterminism.

In short, given my theory of value, it seems to follow that none of the incompatibilist freedoms is valuable. If this is correct, then the first part of the freedom component of evaluative compatibilism – that none of the incompatibilist kinds of freedom that have been proposed as (parts of) accounts of free will is valuable – would be on firm ground. In order to establish the freedom component I will still have to argue that some of the compatibilist kinds of freedom that have been proposed as (parts of) accounts of free will are valuable. I will take up that task in the next section.

3.2 Evaluating the Compatibilist Freedoms

To my knowledge, there are ten prominent compatibilist kinds of freedom that have been proposed as (parts of) accounts of free will. First, it has been argued that free will requires being a persisting substance (an agent) that possesses an ontologically primitive power of ***agent-causation*** that does not consist in and cannot be reduced to causation by events: an agent acts freely if he either exercises that power, or performs an action that is derived from an exercise of that power. If all that you value are positive experiences, the avoidance of negative experiences and the exercise, development and expression of your deliberative capacities in order to perfect your responsiveness to reasons, then this kind of freedom simply is without value, because agent-causation neither consists in nor leads (with a probability that is higher than that of event-causation) to any of these elements.

The same is true for ***noncausal*** theories of free will, that do not explain free actions in terms of what they are caused by or what causal structure they have, but in terms of an agent's being directed at a certain goal (i.e., in terms of her reasons, intentions or purposes) since the fact that an action can be explained in terms of one's striving to attain a certain end does not guarantee in any way either that is is good, or that it will have good consequences, with any degree of probability. And even if explanations of one's actions in terms of one's purposes are themselves valuable (rather than merely a consequence of something that is valuable) for some reason or other, then one may still wonder whether a noncausal sort of explanation can really do the job (Clarke 2008, § 1.2). If you ask me, speaking of free actions as “intrinsically active events” all sounds rather mysterious.

Third, free will can be conceived of as ***active control***, which requires that certain mental events within oneself (in particular motivational and cognitive ones) are the (nondeviant) proximate cause of certain events involving oneself. On itself, this kind of freedom does not have much value: after all, one may well be motivated to act in a certain way by desires that one does not want to be moved by; by desires that are morally outrageous; by a system of beliefs that is ridden with falsehoods; and so on. In general, there simply is no reason to suppose that active control on its own will have good consequences. It may, however, have *some* instrumental value, because active control seems to be a necessary condition of all the kinds of freedom that *are* truly valuable: for it is impossible to perform an action that is very good without possessing the kind of control that is required for acting.

Fourth, some compatibilists have maintained that free will consists in ***the unencumbered power (ability plus opportunity) to do what one wants to do***. It seems quite plausible to me that this kind of freedom *is* quite valuable in general, because being constrained (either by internal or by external factors) in doing what one wants to do evidently is something that usually brings it about that one has some (or many) negative experiences. It is, however, not *very* valuable (not even close), because it neither specifies that your desires or choices have to express your considered preferences, say, nor that your desires have to cohere with rational (including moral) considerations. So although it is plausible that this kind of freedom usually does have *some* good consequences, there is no reason to suppose that it will have, in and of itself, good consequences overall, for all beings involved.

Next in line are ***compatibilist conceptions of the ability to do otherwise***. There are at least five theories of this sort. First, it has been argued that the ability to do otherwise consists in the ability to do otherwise (no internal or external constraints or impediments would prevent one from doing so) if one would want (desire or choose or try) to do so. Since one is only able to *exercise* this ability in certain *counterfactual* situations – if one's mental states would have been different – its possession will never make a difference in the actual world. Moreover, as with the last kind of freedom, being impeded in performing an action may sometimes be desirable: for instance, if an individual is not psychologically capable of harming other sentient beings, or if it would be made illegal to eat meat, then that would (all things equal) be a good thing. So this kind of freedom is not very valuable. But insofar as not being impeded in one's actions usually contributes both to positive experiences and to the development of one's deliberative capacities, it is somewhat valuable. Second, some have argued that being able to perform an action that one does not actually perform consists in *being disposed* to perform that action in a wide enough range of counterfactual conditions. It seems to me that the same points that applied to the previous kind of freedom apply to this one: if one is disposed to do a certain thing only in certain *counterfactual* circumstances, then that disposition will not make a difference in the *actual* circumstances; and it may well be desirable that one is *not* disposed to perform certain actions even in counterfactual conditions (because that disposition might negatively influence one's future conduct in some way). Even so, it may well have some value to be able to do otherwise in this sense in some cases, e.g., because a disposition to perform some action in certain counterfactual situations could *positively* affect one's future actions as well. That is not much, however. Third, the ability to do otherwise has been argued to consist in *having evolved to do otherwise*, in the sense that evolution has enabled us to avoid certain outcomes and to realize certain others. Being "designed" by evolution (via natural selection) to be able to avoid certain things and to realize certain others *might* certainly be valuable, but there is no reason for thinking that it *will* always be valuable, because evolution is by no means a perfect "designer". It is, for instance, wonderful that we are now able to (more or less) critically converse about what the right thing to do is; but evolution has also facilitated wars and genocide, the massive exploitation of other species for trivial purposes, the harming of our environment, and so on. Moreover, in the cases in which it *is* valuable, it is valuable *not* because one has been "designed" to be able to avoid some outcomes and to realize certain others, but because one has been "designed" to be able avoid *the right kinds* of outcomes (bad ones) and to be able to realize *the right kinds* of consequences (good ones), i.e., because one has been "designed" to possess *a valuable kind* of freedom. If it is valuable at all, therefore, it is valuable not in and of itself, but in virtue of what it (contingently) has (happened to) lead to. Fourth, it can be proposed that the ability to do otherwise consists in *adaptive flexibility*, i.e., in being able to do otherwise in *similar* (as opposed to exactly the same) circumstances (its proponent requires more than this – see above – but those elements are discussed elsewhere). If this sort of flexibility genuinely is adaptive in our current circumstances – i.e., if it will allow its possessor to do otherwise in situations that are *relevantly* dissimilar, i.e., in which it is *beneficial* for her to do otherwise – then presumably its possession will (over time) lead both to positive experiences and to the avoidance of negative experience. If this is so, however, then it can probably be reduced to one or more of the other freedoms (those that specify what "the right kinds" of circumstances are). Moreover, the fact that a certain trait (flexibility) is adaptive for an individual evidently does not imply that it will have good consequences for all sentient beings that are affected by the actions of that individual. So although this kind of freedom does seem to have some value, it does not seem to belong to the truly valuable ones (to be fair, however, my judgment of this kind of freedom might be affected by my not knowing much about it). Finally, it could be argued that the ability to do otherwise (the "avoidability" of an action) consists in performing an action without having one's agency bypassed. Similar to the unencumbered power to do what one wants to do, this kind of freedom seems to be valuable for the individual who has it, because having one's agency bypassed will plausibly (at least in certain circumstances) lead to negative experiences (although it may be important how the concepts of "agency" and "bypassing" are specified, so that, for instance, more or less counts as the bypassing of one's agency). Moreover, exercising one's agency probably is a necessary condition of the exercise of any of the truly valuable freedoms; and that at least gives it some instrumental value. Even so, since this kind of freedom does not specify what *content* the exercise of one's agency has to have in order to count as free (that it has to express one's values, that it has to be rational, or whatever), it is not *very* valuable.

Sixth, some compatibilists have argued that free will requires that one has ***taken responsibility for the source of one's action*** (i.e., for the mechanism(s) that bring(s) it about), i.e., (i) that one sees oneself as capable of influencing the course of the world around one through one's deliberations, choices and actions, (ii) that one sees oneself as an appropriate object of the moral expectations and reactive attitudes of other agents, (iii) that one has good reason for seeing oneself in these ways, and (iv) that one has not come to see oneself in these ways through inappropriate (viciously manipulative) ways. In my view, at least the first three conditions of taking responsibility for the mechanism(s) that cause(s) one's action in this sense quite evidently are valuable to some degree. Not seeing one's deliberations, choices and actions (or whatever they are ontologically dependent on) as affecting the course of the world around one might well inspire an anxious sort of fatalism ("there is nothing I can do"), which would plausibly have detrimental consequences both for the relevant individual and for society at large, and seeing oneself as an agent in this sense seems to be a necessary condition of any of the truly valuable freedoms, so the first of the conditions no doubt matters. And the same is true for the second condition: seeing oneself as an apt target for the moral expectations and at least *some* of the reactive attitudes of other members of the moral community probably both is beneficial to the relevant individual (because seeing oneself in that way might prevent a lot of frustration) and to her society (because it plausibly is beneficial to society that people see themselves as subject to certain moral expectations that they have to meet, and that they endorse the appropriateness of certain reactive attitudes, so that their conduct can be influenced by it in positive ways). Finally, appropriately basing these perceptions on one's evidence is symptomatic, at least, of one's being responsive to one's reasoning – and that matters. I am less sure about the fourth condition, however. Sure, it would be better if the agent would come to acquire reasonable beliefs by exercising and developing her own deliberative capacities; and *knowing* that one is manipulated in certain agency-undermining ways might well be harmful to one's self-esteem, and thereby bring it about that one has negative experiences. Even so, there probably are cases in which it will have better consequences overall to ensure that certain individuals will acquire certain beliefs and attitudes through manipulative means (e.g., through manipulating the brains of psychopaths, rather than giving them lifelong prison sentences, or through manipulating children into developing strong dispositions of kindness). Moreover, "moderate" manipulation is very common: we manipulate our children, to some degree, into developing certain traits and dispositions; we manipulate individuals around us into doing what we want them to do; we are flooded with advertisements and propaganda; and so on. So it seems to me that manipulation is not harmful *in itself*: it is only harmful insofar as it *hinders* the development and expression of valuable capacities (or brings it about that an individual has negative experiences). Insofar as it *encourages* them (and positive experiences), however, there is no need to worry.

A seventh framework of theories explains free will in terms of some sort of ***harmony*** or ***mesh*** between certain structural mental aspects of agents. Given that none of these kinds of freedom requires that the relevant mental elements have a certain *content* (e.g., that they are rational or morally defensible), I should mention at the outset that none of them is *truly* valuable within my theory of value. All the same, many of these kinds of freedom are quite valuable. First, it has been argued that free will consists in being moved to action by desires that one wants to be moved to action by (that one identifies with; i.e., it consists in acting on the will one wants to have), in the sense that the relevant desires are the object of an unopposed volition that one is not inclined to change. Since this freedom, like the other "mesh freedoms", does not require that one does not identify with desires that have bad consequences (either for the agent herself (irrational desires), or for other sentient beings (immoral desires)), it is not truly valuable. Insofar as being moved to action by desires that one identifies with reliably produces (more or less) positive experiences, insofar as acting upon desires that one does *not* identify with often produces negative experiences and insofar as all of the most valuable kinds of freedom presuppose that one identifies with one's effective desires, it plausibly is quite valuable. And in my view, the second and third kinds of "mesh freedom" have to be assessed in *exactly* the same way. These theories respectively state that free will consists in being moved to action by desires that are harmonious with one's values (what one values or cares about); and that free will consists in being moved to action by desires that are embedded in one's larger (more or less long-term) plans that one's life is governed by. A fourth theory, however, which says that free will consists in being moved to action by preferences (well-considered higher-order desires) that are coherent with one's other preferences, has to be assessed somewhat differently, for two reasons. The first reason is that individuals often do not notice that their preferences do not cohere with one another; and, if they do notice it, they often rationalize it away. (For instance, most humans believe that it is a bad thing to bring it about that some being has pain. Even so, a large majority of them supports meat companies (through buying their products) that breed animals in circumstances that quite obviously cause them a lot of pain.) This means that incoherences in one's belief system, contrary to failures of the previous kind of "meshes", often do not bring it about that one has negative experiences. In fact, in some cases, incoherences may even produce positive experiences, e.g., if they make you feel good about yourself in unrealistic ways. So with respect to positive and negative experiences, this kind of freedom is less obviously valuable than the previous "mesh freedoms" – although incoherences can, of course, be quite burdensome once you *do* notice them and genuinely see their force. The second reason, however, works in favor of this freedom: that one's preferences are mutually supportive of each other in some way seems to be more tightly connected to the perfection of one's responsiveness to reasons than the other "mesh freedoms". After all, developing one's deliberative capacities precisely seems to involve increasing the coherence between (among other mental states) one's preferences. Assuming that these reasons roughly cancel each other out, I shall tentatively conclude that this kind of freedom is of roughly the same value as the previous ones. Finally, there is the theory that states that free will consists not in performing an action voluntarily, but in performing an action that is expressive of one's judgment-sensitive attitudes (attitudes that are sensitive to judgment in ideally rational agents), so that it can be properly attributed to oneself. Given that one can be free in this sense when performing an action that is both immoral and irrational, that goes counter to one's well-considered desires and that one cannot even control (i.e., given the fact that performing an action that is expressive of one's judgment-sensitive attitudes does not reliably produce any sorts of good consequences), this kind of freedom no doubt is the least valuable of the "mesh freedoms". If it has any value at all, it is so only for our general practices of the moral evaluation of individuals, because it helps us to assess the relevant individual's moral qualities.

An eight kind of conception of free will conceives of it in terms of ***self-control***, i.e., in terms of the ability to act in accordance with the desires that one wants to be moved to action by, that are harmonious with one's values and/or that fit into one's larger policies in the face of strong desires not to do so and other sorts of temptations and distractions. Since self-control basically is the capacity to consistently realize the first three "mesh freedoms" over time, and since all of those are quite valuable for the individual who has them, self-control definitely is a desirable capacity to have – in fact, a *very* desirable one. If it is combined with a strong receptivity for rational (including moral) considerations, then it is a recipe for pure excellence. Moreover, strong self-control may well be correlated with having well-developed capacities for practical reasoning (that's a guess, though). However, as with the "mesh freedoms", even if having self-control reliably produces many good consequences for the individual who has it, since it does not specify what desires one has to identify with or what values and long-term goals one must have, there is no guarantee that mere self-control will encompass responsiveness to reasons and have good consequences for all sentient beings involved. So it does not belong to the truly valuable freedoms, although it definitely comes quite close.

Ninth, some compatibilists have argued that free will consists in ***responsiveness to reasoning***, which consists in the capacity to recognize that one's beliefs and desires have certain implications (that they give one certain reasons) and to appropriately act in response to that recognition (e.g., by modifying one's motives, or by performing different actions than before). In other words, it consists in being able to critically evaluate one's desires, beliefs and values, and to alter them in light of this reflection. Since one can be able to do this even if one's desires, beliefs and values are irrational and immoral (i.e., since mere responsiveness to reasoning does not require that one's reasoning has a certain content – that one has certain kinds of desires, beliefs and values), this kind of freedom is not one of the most valuable ones. Still, responsiveness to reasoning definitely matters, for at least three reasons. First, since responsiveness to reasoning enables individuals who have it to appreciate and respond to the implications of their beliefs, desires and values, it enables them to realize their desires and values better; and that plausibly produces some positive experiences. Second, critically reflecting on one's beliefs, desires and values is one (content-neutral) way of exercising and developing one's deliberative capacities, so that certainly gives it some importance. And third, and relatedly to the second point, the fact that responsiveness to reasoning is a crucial component (and a necessary condition) of responsiveness to rational considerations gives it some instrumental value.

Finally, some compatibilists define free will in terms of ***responsiveness to reasons*** (i.e., to an appropriate range of rational considerations, including moral ones, with respect to which of several actions one should perform) of an agent or of the psychological processes within an agent that bring about an action. In other words, they define free will in terms of being *receptive* to reasons (coming up with and appreciating reasons and evaluating one's beliefs, desires and values in light of those reasons) and in terms of being *reactive* to reasons (responding or reacting to reasons, regulating one's behavior in light of reasons), i.e., in terms of doing the right thing for the right reasons – reasons that derive from "the True and the Good". Based on my theory of value, this kind of freedom definitely is the most valuable of them all. It is not hard to see why: unlike all the other freedoms that I have evaluated, being responsive to rational (including moral) considerations reliably produces good consequences over time, not only for the individual who is so responsive, but for all sentient beings that are affected by her actions – and the more responsive to reasons an agent is, the better. In fact, besides its enormous instrumental value in consistently bringing about positive experiences and avoiding negative ones, my intuitions have even granted it *intrinsic* value: in my view, perfecting one's responsiveness to rational considerations simply is valuable in itself. So the freedom that is most valuable is the freedom to be receptive to rational and moral considerations and to evaluate and (if necessary) adjust one's beliefs, desires, values and conduct in light of these considerations.

Given all this, my values seem to imply that some of the compatibilist freedoms are indeed valuable – and that one of them is *truly* valuable. If this is indeed so, then that would suffice to establish the second part of the freedom component of evaluative compatibilism – that some of the compatibilist kinds of freedom that have been proposed as (parts of) accounts of free will are valuable. Since I have already argued in favor of the first part of the freedom component, this means that the freedom component of evaluative compatibilism is, given my values, well-grounded. In order to establish evaluative compatibilism, however, I will still have to successfully argue in favor of its responsibility component – the component that states that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all *valuable* responsibilities are compatible with determinism. In the next section, I will strive to establish the first part of this component, that is, I will argue that none of the kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility and that are incompatible with determinism is valuable.

3.3 Evaluating the Incompatibilist Responsibilities

In my view, of all the sorts of responsibilities that I have listed above, only one is not compatible with determinism: moral responsibility in the ***desert-based accountability*** sense. To be precise, this theory says that being morally responsible consists in being an appropriate object of the reactive attitudes (including, for many, praise, credit and rewards and blame, sanctions and punishment), which are emotion-laden evaluative attitudes to the behavior of agents (i.e., it says that it consists in it being appropriate, in certain circumstances, that one is *held to account* for one's conduct); that when and whether an individual is an apt target for the reactive attitudes is determined by moral principles; and that these principles state that an agent, in a certain situation, is a fit candidate for some reactive attitude if and only if she *deserves* it that the relevant attitude is directed at her for non-consequentialist reasons (e.g., because she voluntarily has a bad state of mind, such as a bad will or bad intentions, or because she voluntarily is directed at an end that is morally wrong in some way).

Well, I have been convinced, by an argument from Galen Strawson (1998, § 3; see Haji 2009, 169-181 for a discussion of this argument), that this kind of responsibility not only is incompatible with determinism, but is, in fact, *impossible* (or in any case so implausible that it is hard to tell the difference). Since this is – or so it seems to me – a matter that essentially rests on moral intuitions (namely, intuitions about what desert-entailing moral responsibility requires), and since I have no reason – as far as I know – to believe that consulting my moral intuitions reliably produces true beliefs, I do not strive to establish that it (probably) is *true* that desert-entailing moral responsibility cannot exist. Others have different intuitions about the matter that are as coherent as mine, and so be it. Rather, I will merely try to make a case for what I consider to be highly intuitive, without actually expecting to convince those with opposing intuitions. Strawson's argument roughly goes as follows:

P1. In order to deserve it that a certain reactive attitude (here including praise, blame, etc.) is directed at one for acting in a certain way, one must deserve it that a certain reactive attitude is directed at some crucial factors that bring it about that one acts in the way that one does.

P2. One acts, in a certain situation, in the way that one does, because of the way one is, at least in certain crucial mental respects. (I will omit the latter qualification in the rest of the argument.)

So:

C1. In order to deserve it that a certain reactive attitude is directed at one for acting in a certain way, one must deserve it that a certain reactive attitude is directed at the way one is. (From P1 and P2)

P3. In order to deserve it that a certain reactive attitude is directed at the way one is, one would have to intentionally bring (or have brought) it about that one is the way one is.

P4. An infinite causal chain of mental natures (each one intentionally caused by – or: intentionally brought about in light of – the previous one) is impossible (for humans, at least).

So:

C2. In order to deserve it that a certain reactive attitude is directed at one for having a certain mental nature, it has to be the case either that one has intentionally brought it about that one has that mental nature without being caused to do so by a previous mental nature, or that one has so brought about some previous mental nature that has caused this one. In other words, it has to be the case that one is (at some point) the uncaused (but intentional) cause of one's own mental nature. (From P3 and P4: otherwise, one's nature would eventually have to derive from something other than one's reasons)

But:

P5. Intentionally bringing it about that one is the way one is presupposes that one has a certain nature N-1 in light of which one intentionally brings it about that one now has nature N.

So:

C3. Intentional self-origination is impossible. (From P5)

Therefore:

C4. One cannot deserve it that a certain reactive attitude is directed at the way one is. (From C2 and C3)

Therefore:

C5. One cannot deserve it that a certain reactive attitude is directed at one for acting in a certain way. (From C1 and C4)

Besides pursuing Galen Strawson's “positive” strategy (i.e., arguing that this kind of moral responsibility has a certain requirement that cannot be met), one can also pursue a “negative” strategy (i.e., one can also argue that all possible accounts of this kind of moral responsibility fail, indicating that it is impossible). This negative strategy (not mine either) can be pursued as follows:

A1. Either determinism is true or indeterminism is true.

A2. If one is not even partly responsible*[55] for some fact, and if one is not even partly responsible* for the fact that this fact brings about some further fact, then one is not even partly responsible* for that consequent fact (Haji 2009, 31, 78; McKenna 2009, § 4.1; Sommers 2009, 512, note 3).

Assume that:

B1. Determinism is true.

[55] “Responsibility*” here refers to desert-entailing moral responsibility.

In that case:

B2. One's actions are deterministically caused by facts that one is not even partly morally responsible* for (Haji 2009, 78), i.e., events in the distant past and the laws of nature, because they are beyond one's control (Sommers 2009, 511), one does not have, or ever had, any choice about their existence (Timpe 2006, § 4.a), they are not up to one (Vihvelin 2011, § 5), and one cannot act in such a way that they would not obtain (McKenna 2009, § 4.1).

And:

B3. One is not even partly morally responsible* for the fact that those facts entail one's actions.

Therefore:

C1. One is not even partly morally responsible* for one's actions. (From A2, B1, B2 and B3)

Now, assume that:

D1. Indeterminism is true.

In that case:

D2. One's actions are probabilistically caused by facts that one is not even partly morally responsible* for, i.e., events in the distant past and the laws of nature.

And:

D3. One is not even partly morally responsible* for the fact that those facts probabilistically bring about one's actions.

Moreover:

D4. If it is possible for one to perform more than one action given exactly the same beliefs, desires and reasons, then it is (to some degree at least) a mere matter of luck or chance how one acts.

And:

D5. If it is a matter of chance that one performs a certain action, then one is not responsible* for it.

Therefore:

C2. One is not even partly responsible* for one's actions. (From A2, D1, D2, D3, D4 and D5)

Therefore:

C3. One cannot be even partly responsible* for one's actions. (From A1, C1 and C2)

(See Clarke 2010, 266; Fischer 2007, 323; Haji 2009, 29-34, 78-79; Kane 2011, 10; Levy & McKenna 2009, 103-104; Lippert-Rasmussen 2009, § 6; Loewer 1996, 105; McKenna 2009, § 4.1; McLeod 2008, § 3; Sommers 2009, 511-512; Strawson 1998, § 3; Timpe 2006, § 4.a; Vihvelin 2011, § 5; and Warfield 2005, 629 for (parts of) versions of the "consequence argument" (A2, B1 B2, B3, C1; A2, D1, D2, D3, C2); and see Clarke 2008, § 2.2, § 2.3; Clarke 2010, 271; Haji 2009, 187-193; Hodgson 2011, 58, 71-72; Kane 2011, 19-20; Levy & McKenna 2009, 118-120; Strawson 1998, § 3; and Warfield 2005, 623-624 for versions of the "luck argument" (D4, D5, C2).)

So my intuitions are against the possibility of desert-entailing moral responsibility. In my view, however, there is no reason to worry about this: for given my theory of value, this sort of responsibility is entirely devoid of value. After all, *deserving* it that some reactive attitude (or praise, blame, etc.) is directed at one does not reliably produce positive experiences, it does not lead to the avoidance of negative experiences and it does not contribute to the exercise and development of one's deliberative capacities in order to perfect one's responsiveness to reasons – and that, on my theory of value, is all that matters. Given my values, therefore, it simply is *irrelevant* whether some individual deserves it that some reactive attitude is directed at her: that she, e.g., had a wrong intention, does not in any way entitle us to do wrong to her! As the saying goes, two wrongs do not make a right. Rather, what matters is whether it has good consequences that the reactive attitude is directed at her.

Before closing this section, I would like two make two last remarks. First, it is *not* the case, as some seem to suppose (and as I have previously thought), that valuing desert-based moral responsibility is incompatible with consequentialism, because it is perfectly coherent to endorse a version of consequentialism that assigns intrinsic value to this sort of responsibility, or to something (e.g., justice or fairness) that may be thought to presuppose this kind of responsibility. For instance, according to Lippert-Rasmussen (2009, § 2), Fred Feldman has defended a consequentialist view that incorporates desert by stating 'that a pleasure is more valuable if it is deserved and less valuable, or perhaps even disvaluable, if it is undeserved'. Similar accounts, such as one that states that although an increase in well-being (or loss avoidance) is always valuable, it is *more* valuable when it is deserved (Lippert-Rasmussen also discusses this one), can also be consistently defended. And, second, even my kind of consequentialism is not inconsistent with the value of *some* sorts of desert. For instance, it seems to be coherent to claim that some beings *deserve* it that their interests are taken into account, e.g., because they are capable of having positive and negative experiences and because they are capable of perfecting their responsiveness to reasons – and that is perfectly consistent with my form of consequentialism. In a similar spirit, it could be argued that one deserves it that a certain reactive attitude (etc.) is directed at one precisely when one's interests are taken into account and it is concluded that it has good consequences for everyone involved to target one with the attitude. So even my account can incorporate some kinds of desert: it cannot, however, incorporate the sort of desert that entails that an agent can deserve it that a certain reactive attitude is directed at her *independent of (standard) consequentialist considerations* (e.g., because she performed a certain action knowingly and voluntarily). As I argued above, however, there is no need to worry about that.

In short, given the theory of value and the normative ethical theory that I endorse, the first part of the responsibility component of evaluative compatibilism – that none of the incompatibilist kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility is valuable –

seems to be correct. In the next section, I will argue in favor of the second part of the responsibility component – that is, I will argue that some of the kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility and that are compatible with determinism are valuable. If that is indeed so, then evaluative would seem to have very firm foundations indeed.

3.4 Evaluating the Compatibilist Responsibilities

There are, to my knowledge, six compatibilist theories of moral responsibility. First, some have argued that moral responsibility consists in ***attributability***. On this view, an agent is morally responsible for an item if it expresses her moral nature or character – more precisely, her judgment-sensitive attitudes (attitudes that are sensitive to judgment in ideally rational agents) – so that it can properly be attributed to her; and on that basis, she can properly be morally assessed for it – in particular, the character traits that it reveals can properly be measured against some moral standard. And this does not require that the relevant item is *voluntary* (i.e., that it is within an agent's control, in the sense that she can exercise influence over whether it occurs). Given that this sort of moral responsibility simply is the equivalent with respect to moral responsibility of the last kind of "mesh freedom", it must be evaluated in precisely the same fashion: since being morally responsible in this sense does not guarantee that good consequences are (often) produced (for an item that is immoral, irrational and in tension with an individual's values, larger policies and higher-order desires may well be expressive of her judgment-sensitive attitudes), it is of very little value indeed. The only value that it may have derives from the the fact that it is a necessary condition of the functioning of our responsibility practices: after all, directing a certain reactive attitude (etc.) at an agent for some reason evidently requires that we can morally assess the item we hold her accountable for.

Second, some have thought that moral responsibility consists in ***appraisability***, i.e., that it consists in having a moral standing or record that can be enhanced or diminished by events in one's life – or, in other words, in having a "moral ledger" that can be assigned positive, negative or neutral marks. Importantly, moreover, that one's moral standing is affected by some item requires more than mere attributability: it may, for instance, also require that the agent identifies with or values the relevant item, or that it fits into her larger plans. In virtue of restricting the range of what counts as relevantly expressive of an agent's moral nature in this way, it plausibly is more valuable than the previous kind of moral responsibility. Furthermore, it may well be the case that, in many (though probably not all) cases, when considering how to react to an agent, it is useful to consider whether she genuinely wanted (in the more narrow sense) to, say, perform the action; and that may give it some value. Even so, I do have some serious doubts about the concept of having a moral record or ledger that a person carries with her throughout her life. It may be wondered, after all, what precisely the point is of morally evaluating a person now for what she has done years (or even a month) ago, if she has since gone through importance changes, in the sense that her character traits have changed in such a way that she would never dream of performing (or omitting to perform) the same action that her past self (by lack of a better name) is morally assessed for performing (omitting). For instance, a person that carried my name, had the same parents, the same genes, the same birth date and the same physical continuity with earlier individuals that I have ("my earlier self" in short) ate meat – the worst kind of meat, in fact. Sure, that is a moral tragedy. But what has it got do with *me*? I simply don't see why the fact that my earlier self ate meat is at all relevant for morally assessing *me*. So although the more narrow focus of this kind of responsibility certainly is an improvement over the rather broad scope of the previous sort, it also seems to require quite a demanding sort of personal identity over time that it is, in my view, not very reasonable to believe in – and that isn't valuable, either. It is true that this view does not *have* to presuppose that persons have one moral ledger throughout their life: they could also state, for instance, that individuals have different moral records at different points in time – records that express how high the moral standing of that individual is *at that point in time* (in virtue of the character traits and dispositions to perform certain actions she has at that time). An agent may then be morally responsible for an item because it somehow meshes with her volition or values or more or less long-term plans, say, *at that time*. In my view, this sort of responsibility *is* quite valuable – for the relevant agent at least, and probably also for our moral practices with regard to the evaluation of agents at particular times. But the value of its counterpart – the one that requires strong personal identity over time – is severely diminished by its value-reducing demands.

Third, there is a theory of moral responsibility that says that being responsible for an item consists in being ***answerable*** for that item, which means that expecting the agent to explain or justify (to answer critical questions about), e.g., her choice or action, or (in the absence of a justification) to acknowledge that her choice or action was wrong, is appropriate. Presumably, moreover, *a certain kind* of explanation is required: explanation (or justification) in terms of reasons. It seems to me that this kind of responsibility is definitely valuable, for three reasons. First, being able to explain or justify, e.g., one's choice or action, seems to be a necessary condition of developing one's deliberative capacities in order to perfect one's responsiveness to reasons; and that gives it some instrumental value. Second, that it is appropriate to expect of an agent that she explains or justifies her choice or her conduct plausibly reflects the fact that she is, to some degree at least, both responsive to reasoning and responsive to reasons. It reflects the fact that she is responsive to reasoning because being able to answer critical questions about your behavior – e.g., why you did a certain thing and why it was justified – seems to reflect the fact that you are able to recognize what reasons your beliefs, desires and values give you and to act accordingly – otherwise, there neither would be an explanation (in terms of reasons) of your conduct, nor could you give one. And it reflects the fact that she is (to some degree) responsive to reasons because being able to justify your behavior seems to reflect the fact that you are capable of recognizing rational (including moral) considerations and of acting in accordance with them. And that it reflects these facts probably adds to its value. Finally, and probably most importantly, this sort of moral responsibility facilitates certain valuable social practices. For instance, the practice of expecting each other to explain or justify (to answer critical questions about), e.g., one's choices or actions probably encourages the development of one's deliberative capacities, because their minimal development is required for satisfying these expectations – not their perfection, but still. And, as Pereboom (2011, 408) points out, the practice of engaging in critical interactions about whether one's actions were morally right, how one could improve one's conduct, motives and character in the future, and so on, plausibly contributes a lot to the moral improvement of oneself and others. Of course, this improvement will often be rather unimpressive by the standards of my ethical theories – as long as they are not more immoral than what counts as normal in their society (or their section of society), most people will feel no need to change; and the fact that people often come up with a rationalizations rather than genuine explanations or justifications obviously does not help either. Still, the practice plausibly is essential to the moral functioning of our society, and that gives it much importance. In short, this kind of responsibility definitely is quite valuable, but it falls short of *true* value: it simply demands too little to be worthy of that status.

Finally, some have thought that moral responsibility consists in ***accountability***, i.e., in being an apt target for the reactive attitudes (in it being appropriate that one is *held to account* for some items, in particular one's choices and actions), such as admiration, forgiveness, guilt, hatred, love, moral sadness, respect, indignation and resentment. Within this general framework fit at least three compatibilist sorts of interpretations of what "aptness" or "appropriateness" consists in here.

First, some have argued that whether it is appropriate to direct a certain reactive attitude at an individual in certain circumstances is determined not by the nature of that individual, his action and the circumstances in which the action takes place. Rather, whether and when an individual is morally responsible in this sense is socially constructed by the members of the moral community: it simply derives from the normative attitudes and practices that they endorse. To a large extent (excepting a small minority of critical individuals), of course, this is how it actually goes. In order to determine whether some individual is morally responsible, most of us do not consult moral principles of some sort: rather, we simply do what our parents have taught us to do, with some small modifications here and there; we do what people around us do – in short, we do what it is common in our culture, or (section of) society, to do, and thereby follow the normative considerations that happen to be widespread in our social environment. It is easy to be cynical about this, but in my view that isn't warranted. To be sure, our current practices are *definitely* not ideal (judged by my values). Still, given the pivotal role that our responsibility practices play in educating children and in guiding our behavior, and given the fact that it is plausible that selective social processes ensure that they are at least to some degree beneficial in their social contexts, they are quite valuable nevertheless. Moreover, insofar as it is preposterous to expect that a large majority will ever (in the foreseeable future at least) become motivated to rigorously examine their moral convictions, if there will be any major further improvements (in comparison to their past shape, they have of course already improved a lot) in the responsibility practices of our culture, then they will probably result not from a rational convergence on a certain ethical theory, but from complex social processes that it is difficult to even become aware of. And that plausibly gives this sort responsibility some instrumental value. However, its value should not be exaggerated either: improvement usually takes a lot of time; in the wrong sorts of circumstances, improvements can be reversed; and nothing in this interpretation excludes the validity of responsibility practices that have disastrous consequences for all sentient beings that are affected by it. So although this sort of interpretation certainly can be (and probably often is) quite beneficial, it is far too adaptable and normatively undemanding to be highly valuable.

The other two compatibilist interpretations of what "appropriateness" means here are *moral* interpretations, that is, they state that even if the *actuality* of certain responsibility practices is not determined by moral principles (as seems difficult to dispute), their *normative validity* or *rightness* is so determined. So whether a certain individual is an appropriate candidate for certain (or all) reactive attitudes, and whether it is appropriate, in a certain situation, that a certain reactive attitude is directed at a certain individual, is not socially constructed, but determined by moral standards of rightness. The first of these interpretations distinguishes two elements: the legitimacy of the practice of directing one's reactive attitudes at each other in general; and the fairness of directing a certain reactive attitude at a certain individual in certain circumstances. It then states that the general practice is morally legitimate in virtue of the contribution it makes to fostering certain virtuous character traits by informing individuals about and enforcing certain ethical rules and ideals. But it also emphasizes that the practice has to take into account considerations of fairness, so that it is only right to direct one's reactive attitudes (perhaps including praise, blame, and so on) at certain kinds of individuals in certain kinds of circumstances: for instance, it may be required that the individual is sufficiently developed and sane and performed the relevant action, say, voluntarily and knowingly. Judged by my theory of value, this sort of responsibility definitely has a lot of value to it: the development of virtuous character traits – as I have defined them – has intrinsic value on my theory; and their perfection also contributes a lot to the reliable production of positive experiences and the avoidance of negative experiences (because those character traits include rational and moral dispositions). It is true that many accounts of the virtuous life can be proposed that diverge from mine, and their value will of course differ from my account, but I'll leave that complication aside here. It seems to me that the fairness component of the theory does not mesh well with my ethical theories, however. Insofar as it has good consequences only to direct a certain reactive attitude at an individual if she performed the action (to some degree perhaps) voluntarily and knowingly, say – and this may well be so – my theories will imply that it is right to do so. But, importantly, they will not have this implication *for reasons of fairness*, because they simply do not grant intrinsic value to fairness. Finally, note that this theory does not mention the importance of positive experiences and the avoidance of negative experiences at any point, which obviously goes counter to the inclinations of my theories. So even though this sort of responsibility has quite a lot of value, it does not manage to attain *true* value.

The second moral interpretation of what "appropriateness" consists in may also fruitfully be interpreted as distinguishing between two elements: the rightness of the practice of holding each other accountable for each other's behavior in general; and the rightness of directing a certain reactive attitude at a certain individual in a certain situation. On this view, the general practice is morally right because it reliably produces (sufficiently) good (expectable) consequences: first, because directing reactive attitudes (and their more retributive counterparts) at others is an excellent instrument for encouraging or discouraging certain choices, actions and character traits; second, because we are not capable of abandoning all reactive attitudes; and, third, because abandoning all reactive attitudes would be irrational in light of the fact that they are essential to interpersonal relationships. And the account states that (at least in usual situations) all individuals that are capable of determining which of several actions they have good reason to do (which one serves their interests), so that they are capable of changing their conduct in response to sanctions and rewards, are appropriate objects of the reactive attitudes; that individuals who do not perform an action voluntarily and/or knowingly are (at least in many cases) not appropriate objects of the reactive attitudes; and that we should (probably) strive to use the "positive" reactive attitudes *more* than the negative ones, that we should cut down on blame, hatred, indignation and resentment and that we should strive to reform our punishment practices and sanctions in order to improve the value of their consequences – and in all these cases, "appropriateness" is determined by the (foreseeable) value that is produced by adopting a certain course of action. Unsurprisingly, given the fact that this view meshes perfectly with my ethical theories, I do believe that this sort of responsibility is truly valuable: if it would be endorsed by many in our society – it is not, of course – and if more research would be done in order to improve the effectiveness of directing one's reactive attitudes (etc.) at individuals, then the consequences would probably be fantastic. Of all the responsibilities, therefore, this one no doubt is the most valuable.

3.5 Conclusion

In short, judged by my ethical theories, none of the incompatibilist kinds of freedom that have been proposed as (parts of) accounts of free will is valuable; some of the compatibilist kinds of freedom that have been proposed as (parts of) accounts of free will are valuable (and one of them is truly valuable); none of the incompatibilist kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility is valuable; and some of the compatibilist kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility are valuable (and one of them is truly valuable). If this is indeed so – that is, if I have not made some important misevaluations somewhere – then evaluative compatibilism – the thesis that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms are compatible with determinism (the freedom component); and that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities are compatible with determinism (the responsibility component) – would seem to be correct – at least within my ethical theories.

Before proceeding to the next chapter, however, I would like to make two final notes. First, given that all of my evaluations are rather sketchy and tentative, no doubt some of them are wrong in the details. Even so, I am pretty confident about the rough outlines of my evaluations, in the sense that I don't expect to be convinced that some of the incompatibilist freedoms and responsibilities are valuable, or that none of compatibilist freedoms and responsibilities is valuable. Since that is all that evaluative compatibilism needs, the sketchiness of my evaluations probably does not endanger the firmness of the foundations of evaluative compatibilism. And, second, despite the fact that I have argued in favor of evaluative compatibilism on the basis of a very specific theory of value and normative ethical theory, it seems plausible to me that many ethical theories have the implication that evaluative compatibilism is correct. So long as you does not assign intrinsic value to any of the incompatibilist freedoms or responsibilities that I have argued to be devoid of value (e.g., absolutist freedom, or desert-entailing moral responsibility), or to something that presupposes them (e.g., some kinds of justice), and so long as you do not endorse a theory of rightness that requires that some of the incompatibilist freedoms or responsibilities exist (e.g., one that states that an action can only be right or wrong if its performer has the ability to do otherwise given the past and the laws of nature), then you endorse a position that entails that evaluative compatibilism is correct. And – as it is also important to emphasize – these positions do *not* have to be consequentialist ones: it is perfectly consistent to argue that actions are wrong in virtue of their intrinsic natures (deontological ethics), or in virtue of the fact that they either are or would be performed by an individual with a virtuous character (virtue ethics), and still maintain that evaluative compatibilism is correct. So despite my rather narrow focus, evaluative compatibilism is consistent with a wide range of ethical views.

Now that I have argued in favor of evaluative compatibilism, and made the above two notes, I will proceed to argue in favor of the second main thesis that I will strive to establish in this thesis – critical evaluative optimism. Doing so will be the topic of my next chapter.

Chapter 4: In Defence of Critical Evaluative Optimism

In this chapter, I will argue in favor of my second main thesis – critical evaluative optimism, which is the thesis that, of all the kinds of freedom that have been proposed as (parts of) accounts of free will, all valuable freedoms exist in the actual world, and that, of all the kinds of responsibility that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities exist in the actual world (the evaluative optimism component); but that scientific and everyday observations do show that we have some valuable freedoms and responsibilities to a lesser degree than would be ideal (the critical component). Before taking up that task, however, it will probably be useful to provide a brief overview of which freedoms and responsibilities are, in my view, valuable:

The valuable freedoms	**Their value**
Active control	Very low
Being disposed to perform an action in a wide enough range of counterfactual conditions (the second compatibilist conception of the ability to do otherwise)	Very low
Having evolved to do otherwise (the third compatibilist conception of the ability to do otherwise)	Very low
Performing an action that is expressive of one's judgment-sensitive attitudes (the last kind of "mesh freedom")	Very low
Being able to do otherwise (no internal or external impediments would prevent one from doing so) if one would want to do so (the first compatibilist conception of the ability to do otherwise)	Low
The unencumbered power (ability plus opportunity) to do what one wants to do	Medium
Adaptive flexibility	Medium
Performing an action without having one's agency bypassed	Medium
Taking responsibility for the source of one's action	Medium
Being moved to action by desires that are the object of an unopposed volition that one is not inclined to change (the first kind of "mesh freedom")	Medium plus
Being moved to action by desires that are harmonious with one's values (the second kind of "mesh freedom")	Medium plus
Being moved to action by desires that are embedded in one's larger plans (the third kind of "mesh freedom")	Medium plus
Being moved to action by preferences that are coherent with one's other preferences (the fourth kind of "mesh freedom")	Medium plus
Responsiveness to reasoning	Medium plus
Self-control	High
Responsiveness to reasons	Very high

The valuable responsibilities	**Their value**
Performing an action that is expressive of one's moral nature or character (more precisely, one's judgment-sensitive attitudes) (“attributability”)	Very low
Having a moral standing or record that can be enhanced or diminished by events in one's life – or, in other words, having a “moral ledger” that can be assigned positive, negative or neutral marks (“appraisability”)	Very low or medium plus[56]
Whether an individual is accountable in certain circumstances is determined by (the normative attitudes and practices of) the members of the moral community (the first kind of “accountability” responsibility)	Medium
Being answerable for an item, in the sense that expecting one to explain or justify (to answer critical questions about), e.g., one's choice or action, or (in the absence of a justification) to acknowledge wrongdoing, is apt (“answerability”)	Medium plus
Our general accountability practices are legitimate in virtue of their contributing to fostering virtuous character traits; but they have to be constrained by considerations of fairness (the second kind of moral “accountability” responsibility)	High
Our general accountability practices are legitimate in virtue of their good consequences; and whether it is right to hold a certain individual accountable for a certain item in certain circumstances must be determined on consequentialist grounds as well (the third kind of moral “accountability” responsibility)	Very high

Now that I have provided a brief overview of all the valuable freedoms and responsibilities, and of their respective value (but again, there is definitely room for improvement in the details here) I can proceed to state in more detail what I will do in this chapter. In section 4.1, I will rather briefly argue both in favor of the evaluative optimism component and in favor of the critical component on the basis of everyday observations of the world around us and the behavior of the individuals in it. In section 4.2, then, I will discuss four challenges that many have taken to threaten not only the degree to which we have some valuable freedoms and responsibilities, but even their very existence. I will argue that none of these challenges threatens the existence of any of the valuable freedoms and responsibilities, thus reaffirming the evaluative optimism component. I will also argue, however, that the last of the challenges that I will discuss does genuinely threaten the degree to which we have some (very) valuable freedoms in particular, thus reaffirming the critical component. If all this is indeed so – given the theory of value and the normative ethical theory that I endorse, of course – then critical evaluative compatibilism would seem to be correct. Finally, I will argue, in section 4.3, that the critical component should not stem us *too* pessimistic, because things are improving.

[56] It depends on how it is interpreted. For more information, see its evaluation.

4.1 An Everyday Defence of Critical Evaluative Optimism

4.1.1 An Everyday Defence of the Evaluative Optimism Component

In this section, I will briefly provide a defence of the evaluative optimism component on the basis of everyday observation that are accessible to everyone without knowing any scientific studies and without having any detailed knowledge of philosophical positions and arguments. It goes as follows: our everyday experience of the world simply makes it *obvious* that all of the freedoms and responsibilities that I have argued to be valuable indeed exist in the actual world (that is why they have been proposed as compatibilist accounts of free will or moral responsibility in the first place, of course). Look around (or within) you! It *continuously* happens that individuals are moved to action by desires that are harmonious with their values; that they act with self-control; that they act in a certain way in response to rational (including moral) considerations; that they successfully explain and justify their behavior in response to critical questions; that holding individuals accountable for their behavior produces good consequences; and so on. In short, it simply seems to be intuitively and empirically *obvious* that all of the above mentioned freedoms and responsibilities exist in our world. Evidently, this is not a knock-down argument: it can certainly be challenged. But it *does*, at least, firmly shift the burden of proof on those arguing that some of them do not exist in the actual world.

4.1.2 An Everyday Defence of the Critical Component

I have an everyday argument for the critical component as well – and it precisely mirrors my argument for the evaluative optimism component. It states that it simply is empirically and intuitively *obvious* that we do not possess some (very) valuable freedoms in particular, but also some (very) valuable responsibilities, to the degree that would be desirable – in fact, not even close. Once again: look around (or within) you! Many humans are (to some degree and in some respects) mentally ill; being (more or less) coerced to do something you do not want to do is not uncommon; our agency is sometimes bypassed; people are often moved to action by desires that go counter to their higher-order desires, their values and their larger policies; their preferences are often blatantly incoherent; agents often fail to see that their beliefs, desires and values have certain implications, even when confronted with contrary evidence (e.g., that eating meat and not significantly reducing their greenhouse gas emissions are wrong); they often do not have much self-control (why else are there so many overweight, smoking, drinking and lazy individuals?); our responsiveness to reasons certainly is not bad in comparison to its past shape, but it's not great either (a large majority of humans still is superstitious in one way or other; most humans in the West spend far more time satisfying their meaningless immediate desires than perfecting their deliberative capacities; most of them often do not even do what it is in their long-term best interest to do; beyond a small bubble of dear individuals, most people hardly give other sentient beings a second of thought; and so on); we often do not respond appropriately – as I have defined it – to the reactive attitudes (etc.) of others (e.g., we may take praise as a reason to take some rest, even if we do not need it, we may be driven to apathy and cynicism by moral sadness, or we may get angry at someone for criticizing us rather seeing it as an opportunity for improvement); we are far too prone to negative reactive attitudes; we are not as good at forgiving others as we should be – and the list goes on and on and on. In short, as with the previous component, our everyday experience of the world simply renders it *obvious* that the degree to which we have many valuable freedoms and responsibilities, and all of the most valuable ones, is not, to put it mildly, ideal. If anything is beyond the realm of doubt, after all, it is that humans are, in many ways, often silly and laughable beings – and destructive and vicious ones, too. So the foundations of the critical component seem to be not only very firm – but even unshakable.

In short, it seems to me that everyday observations in conjunction with my ethical theories give us good reason to believe that both the evaluative optimism component and the critical component are correct. It has, however, been widely argued that philosophical arguments and scientific experiments give us reason to think that the evaluative optimism component is false – i.e., that some valuable freedoms and responsibilities do not exist – and that an even more critical attitude is warranted than everyday observations suggest. In the next section, I will discuss four challenges of this sort.

4.2 Four Challenges to Optimism about Free Will

4.2.1 The Challenge from Eliminative Materialism

The first challenge that I will discuss largely is a philosophical one, although it certainly can be informed by scientific considerations – it is the challenge posed by “eliminative materialists”. They argue, as Ramsey (2007, introduction) puts it, 'that our ordinary, common-sense understanding of the mind [folk psychology] is deeply wrong and that some or all of the mental states posited by common-sense [e.g., beliefs, desires and sensations] do not actually exist'. The argument for eliminative materialism that in my view is most convincing (see Churchland 1989, 6-9; Horgan & Woodward 1985, 199-200; Ramsey 2007, § 2.1, § 2.2, § 3.1, § 4.2) roughly proceeds as follows. Assume that physicalism is the thesis that everything either is, or is ontologically dependent on the, physical (Stoljar 2009, introduction, § 3); and assume that *reductive* physicalism is the conjunction of physicalism and the thesis that complex physical entities reduce to collections of simpler physical entities (e.g., elementary particles, or strings, or whatever) (Meijsing 2012) (I will leave it open what “reduction” precisely consists in here). Given these assumptions, the following argument can be given:

P1. Reductive physicalism is true.

P2. Folk psychological entities are certain kinds of mental states: they have certain 'requirements that any state or structure must meet to qualify as a mental state of that sort' (Ramsey 2007, § 4.2).

P3. (Some of) those kinds of mental states cannot be reduced to or identified with lower-order (collections of) physical entities, in particular neurological events or processes (Ramsey 2007, § 2.2).

Therefore:

C1. (Some) folk psychological entities do not exist. (From P1, P2 and P3)

Granting (for the sake of the argument) P1 and P2 to the eliminative materialist, what can be said in defence of P3? Focusing on general arguments, there are at least four arguments. First, it has been argued that there are a lot of mental phenomena that folk psychology cannot explain; and this indicates that it is at best a partial description of 'a deeper and more complex reality' (Churchland 1989, 7); and that means that at least *some* folk psychological entities cannot be reduced (Churchland 1989, 6-7; Horgan & Woodward 1985, 199; Meijsing 2012; Ramsey 2007, § 3.1). Second, the argument has been given that, since most other folk entities have been eliminated, since many phenomena that were previously explained in intentional terms are now explained in physical terms, since both 'the content and the success' (Churchland 1998, 8) of folk psychology have barely progressed, if at all, over the last two or three thousand years, and since 'folk psychology concerns a subject that is far more complex and difficult than any past folk theory' (Ramsey 2007, § 3.1), it is probable that folk psychological entities cannot be reduced to physical entities (Churchland 1998, 7-8; Horgan & Woodward 1985, 199; Meijsing 2012; Ramsey 2007, § 3.1). Third, it has been argued that the folk psychological categories seem to be incommensurable with the categories of neuroscience in particular, which indicates that the former cannot be reduced to the latter (Churchland 1998, 8-9; Horgan & Woodward 1985, 199-200; Meijsing 2010). Finally, some have argued that the fact that folk psychology has not come into existence as a result of our striving to develop a theory of our mental life that coheres well with what the natural sciences (neuroscience in particular) have to say about it means that reduction of the former to the latter is unlikely (Horgan & Woodward 1985, 197).

Assume that all the premises hold. In that case, one may still wonder whether there even *is* such a thing as, say, a “folk belief” or “folk desire” – one may wonder, that is, whether ordinary people genuinely have both somewhat detailed and very similar folk conceptions of what certain kinds of mental states consists in. It may equally well be the case that they only have very rough and flexible conceptions of what certain kinds of mental states are, conceptions that are tentatively formed by their innate capacities and tendencies to interpret the behavior of others (and themselves) in certain ways, the use of the words that express those conceptions in their linguistic community and their own experiences of those states – indeed, conceptions that are so flexible that it is very probable that they (or their "hard core", if you will) can be reduced to neuroscientific ones. For instance, they may not have a precise understanding of what "pain" consists in, but simply think that it is, you know, that experience you have when you burn your hands or fall from a stairs or are dumped by your girlfriend – and there is no good reason (as far as I know) to believe that pain *in that sense* does not exist. It exists alright. Moreover, it may well be the case that some of the complicated accounts of certain kinds of mental states that have been offered by philosophers do not exist (cannot be physically reduced); but that others do – and that these are all we need.

But let us grant the eliminative materialist, for the sake of the argument, that ordinary people indeed have somewhat precise and very similar understandings of what certain kinds of mental states consist in; and grant her that some (or many) of these conceptions fail to refer. Would this threaten the evaluative optimism component? Well, no, I don't think it would – and most eliminative materialists probably agree with me on this point. I have three arguments in favor of this conclusion. To begin with, note that many of the valuable freedoms and responsibilities probably are not threatened by eliminative materialism anyway. Rationality and moral excellence? Not a chance. Self-control? Nope. Being moved to action by motivational states of some sort that are harmonious with what one values? Not likely. Being capable of being steered in the right direction by our accountability practices? Of course not. And that at least heavily reduces the threat that eliminate materialism poses.

Second, imagine that it is the case that there is, for example, no such thing as "folk pain": it may, e.g., be the case that it has certain properties, such as infallibility and intrinsic awfulness, that come apart in practice, or that it is far more complex than folk psychology supposes it to be (Ramsey 2007, § 3.3). In response to that, I would simply shrug my shoulders. So "folk pain" doesn't exist. Big deal. Touching a hot stove, or spending too much time behind a computer screen, or being insulted by someone, and so on, still brings it about (in usual cases at least) that I have an awful experience of some sort that I desperately want to avoid, and that really is all that is necessary in order to establish that pain has (again, at least in usual cases) disvalue. And the same goes for other kinds of mental states: even if beliefs and desires do not exist in the precise "folk" form that we ascribe to them, so long as something like them *does* exist in some form that maintains what is important in them, there is no need to worry. As I put it above (in note 12) with respect to events, it may be the case that some folk psychological entities do not exist (cannot be reduced to lower-order physical entities): so long as there are some other entities that are existentially equivalent to them, however – that is, so long as there are entities that are equivalent to them with respect to what matters to us in our everyday life and experience, with respect to their value – their nonexistence simply does not constitute a loss. Compare (as Churchland (1989, 22) asks his readers to do) the demise of vitalism, which is the theory that what makes something alive is a "life force" or "vital spirit": that theory used to be very popular, but no scientifically literate individual still believes that it is true. Still, the non-existence of a "vital spirit" certainly does not entail that we are all dead: rather, it only entails that life is not what we thought it was – it is far more complex, and perhaps somewhat less poetic, but it is not less valuable than a "life force". (The same goes for love – it's nothing supernatural – creativity – it's no divine inspiration – the will – it's not a transcendent faculty – and so on.) In my view, *if* eliminative materialism is indeed correct, then it is quite likely that this will occur for some of the folk psychological entities as well: some of them may not exist, but other (and similar) mental states of roughly equal value do exist, and for that reason nothing is lost. Of course, this is an empirical prediction, so it could turn out to be false: it *could* turn out, that is, that whatever it is that makes some freedom or responsibility valuable does not exist. But that is not the point: the point is that, *if* that threat genuinely is real, then it will not come from eliminative materialism as I have defined it as such, because that only speaks of folk psychological entities, and there is no principled reason for believing that eliminating *them* would be a tragedy. In itself, therefore, eliminative materialism is not worrisome. Rather, if the threat is real, then it will, at most, come from an entirely new area of land that has not yet been explored *on top of* eliminative materialism – territory that indicates that neither some folk psychological entity *nor* a valuable substitute exists. It might certainly be there; but, in any case, speculating that it is there counter to all evidence evidently is of no use at all.

Finally, note that it actually is not important, within my ethical theories, what it is that produces our positive and negative experiences and our rational and moral behavior: rather, what matters is *that* they are reliably produced. Further, note that positive and negative experiences and rational and moral behavior do quite often seem to occur – so, whatever it is, there must be *something* that produces those valuable items. Now, assume that our current conceptual mental territory truly will turn into a battlefield – there are casualties everywhere, and not much of our current conceptual framework survives. Even then, whatever it is that produces the valuable items must survive the slaughter, because it is evident both that the valuable items do exist, and that there must be something that produces them. And if whatever it is that produces the valuable elements survives, then they must survive as well. So it seems that even the widespread nonexistence of entities *similar to* the folk psychological entities does not *have* to have bad consequences: what matters simply is not that consequences are produced by *particular sorts* of entities, but *that* they are produced by some entities, and reliably so; and neither of these hypotheses is threatened in any way by eliminative materialism.

In my view, therefore, the challenge from eliminative materialism – one that is almost univocally believed to be absolutely terrible if true – probably fails. So let's move on to the next one.

4.2.2 The Challenge from Mechanism

The second challenge, which is posed (mainly, but certainly not exclusively) by contemporary neuroscience in particular is that of what I will call, following Nahmias (2010, 346), “mechanism”. That is the view that mental phenomena can be causally explained in terms of the nature and organization of and the interactions between the neurobiological parts on which they are, at least, ontologically dependent (I owe this way of conceiving of and describing mechanism to Nahmias (2010, 346), although my definition differs from his in some details). To be fair, scientists are not yet sure whether mechanism is true – it might still turn out to be false. Neuroscience *is*, however, providing inductive support for its truth, by offering evidence 'that human decision-making and behavior can be explained and predicted in terms of underlying mechanisms' (347). Finally, it should be emphasized that mechanism – contrary to the use of some scientists – is *not* (a version of) causal determinism as I have defined it, because not only is the truth of mechanism compatible with the falsehood of determinism (e.g., 'if any of the component parts in a mechanistic system interact in *indeterministic* ways' (ibid.)), but the truth of determinism is compatible with the falsehood of mechanism as well (e.g., if it is impossible to causally explain psychological processes in terms of neurobiological processes, even though both of these processes are deterministic (ibid.)). And neither are neuroscientists (at least at this point) capable of providing solid evidence for determinism, because 'most of their discoveries involve statistical correlations that are compatible with indeterminism' (346).

So that's mechanism! But is it threatening? Well, no – not even a little bit. Why not? Well, as with determinism, it simply does not entail that any of the valuable freedoms or responsibilities does not exist: their existence is perfectly consistent with the truth of mechanism. For instance, surely the unencumbered power to do what one wants to do is compatible with that power's being causally explainable in terms of underlying mechanisms; surely one can possess self-control even if one only has it in virtue of having certain neurobiological powers; surely one can respond to rational considerations even if one's doing so is ontologically dependent on certain neurobiological mechanisms; and so on. All valuable freedoms and responsibilities are – I think – compatible with the falsehood of mechanism as well, however. It could, for example, be the case that one's being responsive to reasons is ontologically emergent from certain physical mechanisms, in the sense that it is ontologically dependent on them but cannot be causally explained in terms of (and is not predictable from) those mechanisms; or it could be the case (though it isn't likely) that one's choosing to perform a certain action involves (let us suppose here) nonphysical powers of substance-causation. So it seems to be the case, in short, that the truth of mechanism neither threatens any of the valuable freedoms or responsibilities, nor increases the probability that some (or many) of them exist. Rather, mechanism is perfectly neutral with respect to their existence. So it's uninteresting for my purposes. Next.

4.2.3 The Challenge from Scientific Epiphenomenalism

Like the challenge from mechanism, the challenge from scientific epiphenomenalism arises from work in neuroscience – and from work in social psychology, too. Before being able to state what scientific epiphenomenalism is, however, I first need to introduce some concepts. Following Mele (2011, 501), I shall understand "deciding to A" (i.e., deciding to do something) as 'a momentary action of forming an intention to A'; and I shall assume that intentions can be acquired or formed even if no decision has been made about what to do. Moreover, still following Mele, I shall distinguish between "proximal" decisions and intentions, which are decisions and intentions 'about things to do straightaway' (ibid.), and "distal" decisions and intentions, which are decisions and intentions 'about things to do later' (ibid.). Furthermore, I shall distinguish (yet again following Mele) "wanting (or having an urge) to A" from "deciding to A," because people can 'want to do things that they decide not to do' (ibid.) and because they often decide which of several incompatible actions they will perform, and from "intending to A," which I take to be more 'tightly connected' (ibid.) to doing A than merely wanting to do A. Now, finally, I shall, again following Mele (508), distinguish "philosophical" from "scientific" epiphenomenalism. Philosophical epiphenomenalism is '[t]he thesis that although all mental events are caused by physical events, no mental events are among the causes of any physical events' (ibid.). Scientific epiphenomenalism, on the other hand, is a much more narrow view. It is the thesis that 'neither [conscious] proximal intentions* ['a collection composed of [conscious] proximal intentions, their acquisition and persistence' (ibid.)] nor their physical correlates are among the causes of (…) corresponding overt [i.e., bodily] intentional actions' (ibid.) (I have added the "conscious" bit). In the vocabulary of Nahmias (2010, 348), it is the thesis that neither 'our conscious experience of deciding or intending an action' nor its neural correlates are 'causally responsible for producing that action; instead, distinct, non-conscious processes or modules' (i.e., processes that are neither a conscious experience nor the neural correlates of such an experience) cause both the action and the above mentioned conscious experience, so 'that the processes underlying our [conscious] choices and intentions do *not* in fact cause our actions'. According to Mele (2011, 499-501, 508-510), Nahmias (2010, 348-350) and Walter (2011, 518-519, 522) one prominent neuroscientist, Benjamin Libet, and one prominent psychologist, Daniel Wegner, each have given one very influential argument in support of this view. Libet's argument goes as follows:

P1. Experiments indicate that the conscious proximal intention* of subjects to initiate a certain muscular motion (e.g., to flex their wrist) is initiated or prepared by an unconscious process – a process that neither is a conscious experience, nor represents the neural correlates of such an experience.

P2. If a movement or action is initiated or prepared by an unconscious process, then it is not caused by a conscious proximal intention* to initiate that movement.

Therefore:

C1. Experiments indicate that scientific epiphenomenalism is true. (From P1 and P2)[57]

Unsurprisingly, the argument has been widely challenged. To begin with, some have objected to P1 by arguing that the design of the relevant experiments suffers from various methodological problems (see, e.g., Edmonds & Warburton 2012; Mele 2011, 502-503; and Walter 2011, 519); and others have interpreted the results in ways that are consistent with a causal role for our conscious proximal intentions* (e.g., by arguing that the relevant unconscious process merely *motivates* rather than *deterministically causes* the muscular motion (Mele 2011, 503), that the muscular motion is caused by an unconsciously initiated urge to move in conjunction with a conditional conscious proximal intention* to initiate movement whenever such an urge is detected, or that the urge represents a signal for the subjects to then decide whether to initiate the muscular motion or not (Mele 2011, 504-505; see also Nahmias 2010, 352 and Walter 2011, 519)). P2 is even more dubious, however. First of all, if it is true, then it can be argued that C1 follows trivially (without the help of P1), because 'I can neither decide what reasons there are, for me, nor can I assign weights to these reasons, nor, finally, can consciousness settle which course of action the balance of these reasons supports. (…) The demand that we exercise conscious will seems to be the demand that we control our controlling. And that demand cannot be fulfilled' (Levy 2005, 73), on pain of an infinite hierarchy of control systems. More importantly, however, P2 is plainly false: that conscious proximal intentions* are initiated or prepared by an unconscious process simply does not imply that they cannot cause movements to happen. Nor is it surprising that our movements are initiated or prepared by non-conscious brain processes. After all, delaying preparation until after an intention has been formed is rather uneconomical. So the first argument – as it stands – almost certainly fails. On to Wegner's argument:

P1. Experiments indicate that 'in some circumstances, people are not conscious of some of their actions' (Mele 2011, 508), i.e., they do not *feel* that they are the ones performing the action; 'in others, people believe they intentionally did things that, in fact, they did not do' (ibid.), i.e., they *do* feel as if they are the ones performing the action; 'and in yet others, people do things “automatically” and for no good reason' (ibid.).[58] (See Mele 2011, 508-509 for some examples of this.)

P2. In these cases, neither conscious proximal intentions* nor their physical correlates are among the causes of corresponding bodily intentional actions.

P3. All actions are caused in basically the same way. (See Mele 2011, 510 for this premise.)

[57] It should be emphasized that Libet was not a pessimist: he actually believed that we *can* consciously *select between* unconsciously initiated movements. Once we become conscious of an unconsciously prepared movement we can, in his view, determine whether it will go through or not, i.e., we have the power to *veto* our unconsciously formed initiatives (Mele 2011, 499-501, 505-506). This, however, is consistent with scientific epiphenomenalism because the exercise of our veto power does not involve the bringing about of a bodily intentional action. Rather, *if* it is exercised, it only involves *not* bringing about a certain action. And if it is not exercised, then it obviously is causally irrelevant.

[58] He also appeals to Libet's argument, but, having discussed that already, I'll leave that aside.

Therefore:

C1. Neither conscious proximal intentions* nor their physical correlates are ever among the causes of corresponding bodily intentional actions. (From P1, P2 and P3)

Moreover:

C2. Conscious proximal intentions* and their physical correlates are unnecessary for producing bodily actions: non-conscious processes can produce bodily actions as well. (From P1 and P2)

P4. If conscious proximal intentions* and their physical correlates are unnecessary for producing bodily actions, then they are irrelevant for producing bodily actions. (See Nahmias 2010, 350.)

Therefore:

C3. Conscious proximal intentions* and their physical correlates are irrelevant for producing bodily actions. In other words: neither conscious proximal intentions* nor their physical correlates are ever among the causes of corresponding bodily intentional actions. (From C2 and P4)

Therefore:

C4. Scientific epiphenomenalism is true.

As far as I know, P1 and P2 are uncontroversial. It could perhaps be argued that something's being an action requires that it is caused by a conscious intention of some sort; if so, then not being conscious of an action rules it out that it's an action. But replacing "action" by, e.g., "bodily motion" or "muscular activity" should solve that problem. It is far less clear, however, that P3 and C2 are correct. First, it has been objected (against P3) that there is no reason for believing, at this point, that all actions are caused in basically the same way: the fact that the experience of having a certain conscious proximal intention* *can* be mistaken in unusual circumstances simply does not imply that it *must* always be mistaken (Mele 2011, 511; Nahmias 2010, 351). Compare visual illusions: the fact that our visual experiences are sometimes mistaken does not entail that they always are (Nahmias 2010, 351). In my view, this objection is correct: perhaps all actions are indeed caused in basically the same way, but simply assuming that this is so will not get us any further. But C2 is harder to dislodge. First, note that P4 is quite plausible: consciousness takes up a lot of energy, so it seems quite reasonable to suppose (for evolutionary reasons) that if it is unnecessary for doing something, then it is irrelevant for doing that thing. Moreover, note that P1 and P2 indeed seem to imply that conscious proximal intentions* and their proximal intentions are unnecessary for producing bodily actions. Still, it seems to me that the argument fails, because it could well be the case that even though the element of conscious proximal intentions* that makes them conscious is causally irrelevant for producing actions, proximal intentions* *that we are conscious of* (or their physical correlates) may still causally produce actions in some cases – it is just that the 'consciousness of the intentions (or the physical correlates of the consciousness) is not involved' (Mele 2011, 511) in producing the actions (according to Mele, this was first observed by Richard Holton). Furthermore, it may also be the case that conscious proximal intentions* are unnecessary for producing some sorts of actions, but not for others. In short, therefore, given that both arguments for scientific epiphenomenalism fail, we have no convincing reason, at this point, to believe that it is correct.

Assume, however, that scientific epiphenomenalism is true – as it could be. How bad would that be? Well, it actually would not be bad at all. First, note that scientific epiphenomenalism is not as unequivocal a thesis as it may initially appear: there are, after all, many notions of consciousness, and scientific epiphenomenalism may be true for some of them, but not for others. Second, if the concept of *unconscious* proximal intentions* (proximal intentions* that an agent is not conscious of) is coherent – and it seems to me that it is – then even if *conscious* proximal intentions* turn out not to be causally efficacious, *unconscious* proximal intentions* may well be (Mele (2011, 511)); and if unconscious proximal intentions have equally good consequences as conscious ones, then there is, for me, no reason to prefer the latter over the former. And introspectively, this notion actually seems rather plausible to me: in my experience, it hardly ever happens that my actions are preceded by conscious proximal intentions to perform them. For instance, when I am engaged in a conversation with someone, I rarely form a conscious intention to utter a particular sentence – I simply *do* it; and the same is true when I ride on my bicycle, when I cook, when I turn the page of a book – really, almost anytime. Still, many of the actions that I perform when conversing with someone, riding on my bicycle, cooking, etc., seem to be excellent examples of intentional actions. Third, given that scientific epiphenomenalism only states that conscious *proximal* intentions are causally irrelevant with respect to producing bodily actions, it leaves it open that conscious *distal* intentions and conscious planning, deliberation, reasoning and monitoring *are* so relevant (Mele 2011, 507, 510, Nahmias 2010, 352-353; Nichols 2008a, lecture 14) – and in my view, these matter a lot more than proximal intentions. If anything, after all, *these* are the mental phenomena that make a long-term difference to how our lives, and the lives of those around us, go: what matters is not that proximal intentions bring about our actions at any tiny little moment in time, but that our actions are in accordance with our distal intentions, with the plans and reasons that we have endorsed after deliberating about them for some time, with our considered desires, and so on. Finally – and most importantly – scientific epiphenomenalism simply does not threaten any of the valuable freedoms or responsibilities, because none of them requires that conscious proximal intentions* or their neural correlates are among the causes of corresponding overt intentional actions. Being moved to action by desires that are harmonious with one's values or larger plans, self-control, responsiveness to reasons – none of them mentions (or presupposes) conscious proximal intentions*. In Levy's (2005, 71) words, 'it doesn't matter, from the point of view of free will, whether we initiate our actions consciously or unconsciously (…) what matters is control, and control need not be conscious' in that sense. Amen.

For the heck of it, let me also briefly evaluate the importance of *philosophical* epiphenomenalism – the view, as I said, that all mental events are caused by physical events, but that no mental events are among the causes of any physical events. Surely *that* would be terrible? Well no, not necessarily. In principle, at least, the consequences in a world in which philosophical epiphenomenalism is true could be equally good, or even better, than those in a world in which it is false, because physical events could be equally reliable producers of good consequences (or more so) as mental events. For instance, it could be the case that our well-considered desires, our deliberation and reasoning and our decisions and distal intentions all do not make a difference to the physical world. Even so, our behavior in the physical world may exemplify self-control, rationality and moral excellence – and if so, then, on my view, nothing at all is lost: given my values, it simply is not important *how* good consequences are produced (unless the process is valuable in itself) but only *that* they are produced. Compare being a passenger in one of the self-driving cars of the future: if they are better at driving than you are, then it simply does not matter that you are not steering the wheel. In light of all this, it is probably fair to say that my previous evaluation was incomplete: most of the freedoms or responsibilities that I have identified as valuable are not valuable *unconditionally*, but only given the assumption that our mental states indeed make a difference to how our lives go. For instance, what really is valuable is not so much that we are *moved to action* by desires that are harmonious with our higher-order desires, values and larger policies, but that we *act in accordance with them*; what matters is not so much that our *psychological processes* respond to rational (including moral) considerations, but that the *processes that bring about our actions* – whatever they are – are responsive to those considerations; and so on. So, on reflection, it seems that my ethical theories make the initial threat that philosophical epiphenomenalism represents dissolve as well. On to the last one.

4.2.4 The Challenge from Irrelevant Unconscious Influences and Confabulation

The last challenge that I will discuss comes from the quarter of social and cognitive psychologists. Many of them have been arguing for quite some time now (i) that many unconscious cognitive and emotional processes as well as contextual features that we do not want to be influenced by heavily influence our actions and decisions, and (ii) that, being unaware of these influences, we "confabulate" that we performed the actions in virtue of having certain plausible reasons. To begin with the first element, some moral psychologists have suggested that we often act on our immediate gut reactions, and then try to *rationalize* (explain or justify) them in conscious deliberation (Nahmias 2010, 353; Sie & Wouters 2010, 127); social psychologists have indicated that unconscious associations and the activation of stereotypes can affect our decision-making processes and behavior (Edmonds

& Warburton 2012b) and that we are often unconsciously influenced 'by situational factors of which we are unaware and whose influence we would not accept, were we aware of them' (Nahmias 2010, 353; see also Sie & Wouters 2010, 127, 130 and Walter 2011, 521); and cognitive psychologists have shown that our decision-making processes 'often rely on all kinds of rules of thumb (…) that lead to systematic reasoning errors ("biases")' (De Regt, Dooremalen, & Schouten 2007, 483) as well as serious errors in gathering, interpreting and remembering information (483-485). In short, contemporary work in psychology seems to show that many of 'our actions do not properly derive from decisions or intentions that we have consciously considered or would accept as our reasons for acting. Rather, [many of, SG] our actions are produced by other factors, and we *rationalize* them after the fact' (Nahmias 2010, 353-354). It might be asked, at this point, how it is even possible that we so blatantly overlook the factors that bring about our actions – and social psychologists have given humility-inspiring answers to that question as well. Noting (besides the points mentioned above) that we sometimes fail to recognize that one or more of our cognitive states causally produces an action and that we sometimes attribute a causal role to cognitive states that are not causally relevant to producing our actions, they argue that we do not directly and infallibly *perceive*, for many, at least, of our mental states and events, the causal processes that translate some of them into behavior, by way of introspection (Sie & Wouters 2010, 126-127; Walter 2011, 521-522)). Rather, we *infer* (as Nisbett and Wilson have argued (Sie & Wouters 2010, 126; Walter 2011, 522)) (not necessarily unreliably) which mental states have caused our behavior by using an implicit theory (i.e., a body of rules and principles) – more precisely, we do it by determining 'which of the usual causes were present at the time of the action' (Walter 2011, 522) (because we did not attend to, and so did not remember, the unusual ones), or by confabulating one if we cannot find a plausible candidate. Or we may (as Wegner has argued (Sie & Wouters 2010, 126; Walter 2011, 522)) infer whether a particular mental state of ours has initiated or originated a particular action in accordance with the principles of priority (the mental state occurred in our minds just before we performed the action), consistency (the thought is consistent with the action) and exclusivity (we do not perceive an plausible alternative candidates for bringing about the action). In any case, both paradigms state that introspecting whether a certain mental state has caused an action is a process not of *perceiving* the causal process, but of *inferring* that process from information that we have after having performed the action.

Assume that the experimental evidence for this challenge holds. If so, then this challenge *does*, in my view, threaten some (very) valuable freedoms and responsibilities – not their existence, but the degree to which we have them. It is not, to be sure, worrying that we do not introspectively perceive the causal processes that convert some of our mental states into actions, but infer their causal role after the fact; and it is not worrying that we are influenced a lot by unconscious emotional, cognitive and situational factors either. What *is* worrying, however, is that our decision-making processes and actions sometimes (perhaps even quite often) are influenced by unconscious processes that we do not *want* to be influenced by (that we think ought to be *irrelevant* to our decisions and actions); and that we cannot *know* (or at least not at all easily) when this is the case because we confabulate that we decided or acted as we did in virtue not of those pernicious influences but on the basis of having certain plausible reasons. For it is not difficult to see that the degree to which we have some (very) valuable freedoms is genuinely threatened by this. It threatens, for instance, the freedom to act in accordance with our volitions, our values and our more or less long-term plans, because irrelevant unconscious influences evidently might steer us in a direction that we would not endorse, were we aware of their influence; it threatens our ability to achieve coherent preferences, because we are disposed to rationalize incoherences away; it threatens our ability to see what the implications of our beliefs, desires and values really are, because irrelevant unconscious influences distort our reasoning processes; and, most importantly of all, it threatens our ability to appreciate and react to rational (including moral) considerations, because we are sometimes moved to action not by good reasons but by vicious unconscious processes even when we *think* that are being responsive to reasons. As (almost) always, one can give a positive ring to the matter, by arguing (i) that we are not passive victims of irrelevant unconscious factors and processes, but can work to diminish their influence (e.g., by "blinding" ourselves to certain kinds of information, or by acquiring knowledge about these pernicious unconscious influences and checking whether we fall prey to them), and (ii) that we can use those factors and processes to our advantage by steering people in the right direction by playing into them (e.g., by "engineering" situations). But it can equally be pointed out, in a more negative take on the matter, that most people simply are not motivated to do (i), that (ii) can be as easily misused as it can be used to a good end, and that both (i) and (ii) are in any case long-term hopes that are not very relevant to current concerns. So, unfortunately, the critical component is convincingly reaffirmed by this challenge. It does not, however, indicate that the evaluative optimism component is incorrect: after all, that our capacities for conscious deliberation are *sometimes* (perhaps pretty often) aimed at explaining and justifying our behavior and at trying to convince others that they should act in certain ways rather than finding out what it is good to do does not imply that they are *always* are so directed – and a lot of our everyday observations count against the latter statement (Nahmias 2010, 353-354). It could, for instance, be the case that a "dual-system approach" is correct, i.e., it could be the case that we have both a reflexive (automatic and impulsive) and a reflective (controlled and deliberate) system, and that the latter can, to some extent, control the former (Walter 2011, 522); or it could be the case that although certain unconscious influences are indeed pernicious, others are benign; or that our decision-making processes work well in some circumstances, and for some purposes, but not in and for others; and so on. In any case, there is no good reason to believe that we are never responsive to our reasoning or to reasons – nor has this radical position been defended, as far as I know, by any psychologist. So despite the strong support that this challenge provides for the critical component, it does not threaten the evaluative optimism component.

In short, although none of the four challenges that I have discussed manages to establish that some of the valuable freedoms and responsibilities does not exist, the last challenge *does* firmly reaffirm that the degree to which we possess some valuable freedoms is not, to put it mildly, great. In the next section, however, I will strive to spray some optimism even over the critical component.

4.3 The Critical Component Is Becoming Less Critical

Even with respect to the critical component, there is, as I said, some room for optimism. In a word: no matter how dire the current situation may be, it is a major improvement over the past situation – and it is improving still. To be sure, this is not – yet – true for all valuable freedoms and responsibilities. For instance, it is not evident that we are more often moved to action by desires that are harmonious with our higher-order desires, values or larger plans than previous generations; we do not seem to get better – certainly not on our own – at recognizing and genuinely remedying (rather than explaining away) incoherences between our preferences; if anything, the continuous presence of delicious food and alcoholic beverages for very little money and the sea of entertainment systems that surround us all day effectively render it impossible, for many people, to consistently exercise self-control over a day; and so on. Even so, one highly valuable freedom definitely is improving – and it is the most valuable of them all: responsiveness to reasons. Sure, most people still eat meat, but a small and growing minority of individuals is vegetarian (and an even smaller minority is vegan), many others opt for meat that is produced in circumstances that are, at least, somewhat better than they could be, conditions in factory farms are slowly improving (at least in Europe), awareness of the moral status of nonhuman animals is increasing – and all this is true even though the welfare of nonhuman animals was not even an issue fifty years ago. Sure, the policies of governments, multinational corporations and banks are often still rather cynical, in the sense that they often are heavily directed at maximizing the satisfaction of their own interests, to the disadvantage of other parties, the environment and the world economy; but government policies often are less cynical than they used to be, multinational corporations increasingly (albeit slowly) choose to be more ethical because not doing so would damage their reputation and ethical alternatives are increasingly easily available. Sure, consumers, governments and corporations do not do even nearly enough in order to combat climate change, but here, too, things are improving at a sluggish pace. We are becoming less sexist, racist and excessively nationalist. Our scope of concern is expanding. We are becoming more scientifically literate. Many people are becoming less superstitious. Science keeps progressing. Highly reliable information is easily accessible virtually everywhere. Wars are decreasing in number and severity. Academic discussion has never been so quick, global and open as it is now. And so on.

Moreover, even if the other valuable freedoms and responsibilities are not improving now, they will probably improve a lot in the future, for three reasons. First, positive psychology – the branch of psychology that studies what makes a life go well, and how to make it better – is just in its infancy now. Once it will have matured, however, it will probably spawn loads of reliable information on how to maximize self-control, how to better see what the implications of one's beliefs, desires and values are, how to avoid being tricked into falsehoods by one's biases, how to foster certain valuable character traits, such as empathy, kindness, high moral sensitivity, intelligence and rationality, and so on. And this information evidently will find its way not only to individuals eager to improve their lives (or at least some aspects of it), but will also guide and influence government policies directed at improving the well-being (happiness, intelligence, rationality, moral excellence, etc.) of the population that they govern. And psychologists will probably also get better at treating, preventing and even curing mental illnesses. Second, there is reason to think, at least, that several scientific and technological developments, such as the development of genetic engineering, increasingly sophisticated robots, nanotechnology and information technology, will cooperate in order to produce humans that have more self-control, are more capable of seeing what the implications of their beliefs, desires and values are, that respond better to the reactive attitudes of others, whose preferences are more coherent, who are more capable of acting on desires that are harmonious with their higher-order desires, values and larger plans and who are more responsive to reasons (more rational and more morally sensitive) (among many other valuable things) (see, e.g., Garreau 2006). Finally, once we have improved ourselves in such a way – i.e., once we will have created 'greater-than-human intelligence' (Garreau 2006, 82) – we might proceed to improve ourselves still further (using that enhanced intelligence); and once we have done that, we might improve ourselves yet further – until we improve ourselves 'at such a rate as to exceed comprehension' (ibid.) – our current comprehension, that is. To be sure, all of these options may never happen. In particular, there is a lot that can go wrong: flesh-eating viruses may wipe out the human species; self-replicating nanobots may devour all energy on Earth; highly intelligent machines may destroy or enslave the human species; genetic engineering may have unintended consequences; climate change and nuclear warfare might destroy our planet; and so on (Garreau 2006, 136-185). Still, assuming that we do not bring about our own destruction, given that we potentially have billions of years to improve ourselves (highly transformative changes may require only thousands or even merely hundreds of years, but anyway) and given that we have already accomplished quite a lot – scientifically, technologically, philosophically, politically, morally – in a few hundred years, their becoming actual is still quite likely in my view.

In short, it seems that the critical component is on firm – even close to unshakable – footing. As I have argued, however, this does not mean, that it should drive us toward pessimism with respect to the human predicament. It should certainly make us critical about the current state of the human species: we simply are not that wonderful a being as some seem to think. But some of us *are* truly wonderful; many others are pretty alright; and, all in all, we've improved a lot and are improving still, even if far more slowly than would be desirable. And that is not only something. It is quite a lot.

4.4 Conclusion

In short, it seems that we have good reason to believe, given my ethical theories at least, that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms exist in our world, that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities exist in the actual world, that the degree to which we possess some valuable freedoms and responsibilities is not ideal and that the degree to which we have some valuable freedoms and responsibilities has at least improved quite a lot and is improving still. And if all of this is indeed so, then critical evaluative optimism would seem to be on firm footing.

Before proceeding to the conclusion of this thesis, however, I would like to emphasize two points that are similar to the two important notes that I have made in the conclusion of chapter 3. First, although I am quite confident about both the evaluative optimism component and the critical component of critical evaluative optimism, I must admit that my conclusions with regard to the latter might be open to modification and improvement in some details. In particular, it might the case that there are important challenges to the degree to which we have some valuable freedoms and responsibilities that I am not familiar with, or that simply have not been discovered yet; and the future might be less rosy than I hope that it will be. If this is so, then more pessimism about the human condition might be warranted than I am inclined to endorse. And, second, like evaluative compatibilism, the force of critical evaluative optimism is not restricted to the particular ethical theories that I have adopted: many other ethical theories – probably less than with regard to evaluative compatibilism (given that the actual world plausibly allows for fewer freedoms and responsibilities than mere determinism), but still – are harmonious with the truth of critical evaluative optimism. So don't let yourself be fooled in believing that the correctness of critical evaluative optimism is restricted to my ethical theories by my rather narrow focus: it certainly was necessary to adopt those ethical theories in order to evaluate the various freedoms and responsibilities that I have distinguished with some precision, but it was not at all required in order to establish either of my main theses. My ethical theories certainly are one way of getting to the conclusions that I have reached. But other roads can be taken.

Conclusion

Assume that both of my main theses are correct. Assume, that is, that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms are compatible with determinism; that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities are compatible with determinism; that, of all the freedoms that have been proposed as (parts of) accounts of free will, all valuable freedoms exist in the actual world; that, of all the responsibilities that have been proposed as (parts of) accounts of moral responsibility, all valuable responsibilities exist in the actual world; and that everyday and scientific observations indicate that we have some valuable freedoms and responsibilities to a lesser degree than would be ideal. What *then*? Well, nothing. Problem solved. Game over. Finished. If you endorse the right kind of normative ethical theory, then the free will problem simply does not have a bite. It does not even tickle. It is a fun exercise, but it certainly is not something that humans should discuss for milennia. We have better things to do. That, in any case, is where I stand at this point. If you don't agree with me, then that is alright. Just acknowledge that you hold onto a different opinion not because you have understood something that any reasonable human being can grasp, something essential and important about the human predicament, but because your ethical views are different from mine. If you acknowledge *that* – that is, if you acknowledge the importance of ethics to the free will debate and that the free will problem is a problem *only* within certain kinds of ethical theories – then I shall be content. That, after all, is all that I could have reasonably hoped to achieve anyway.

Bibliography

Bishop, R. C. (2011). Chaos, Indeterminism, and Free Will. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 84-100). New York: Oxford University Press.

BBC (Producer). (2011) In Our Time. Free Will. Retrieved on August 6, 2012, from http://www.bbc.co.uk/iplayer/episode/b00z5y9z/In_Our_Time_Free_Will/

Brink, D. O. (2007). Some Forms and Limits of Consequentialism. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory* (pp. 380-423). New York: Oxford University Press.

Buss, S. (2008). Personal Autonomy. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved on August 6, from http://plato.stanford.edu/entries/personal-autonomy/

Center for Inquiry – New York, The. (Producer). (2011) The Science and Philosophy of Free Will. Retrieved on August 6, 2012, from http://www.youtube.com/watch?v=4yp3Wr2yKrY

Churchland, P. M. (1989). Eliminative Materialism and the Propositional Attitudes. In P. M. Churchland (Ed.), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (pp. 1-22). Cambridge, Mass.: MIT Press.

Clarke, R. (2008). Incompatibilist (Nondeterministic) Theories of Free Will. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/incompatibilism-theories/

Clarke, R. (2010). Freedom and Responsibility. In J. Skorupski (Ed.), *The Routledge Companion to Ethics* (pp. 290-301). Abingdon: Routledge.

Crisp, R., & Chappell, T. (1998). Utilitarianism. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy* (Vol. 9). London: Routledge.

Deigh, J. (2011). Responsibility. In D. Dolinko & J. Deigh (Eds.), *The Oxford Handbook of Philosophy of Criminal Law* (pp. 194-217). New York: Oxford University Press.

De Regt, H., Dooremalen, H., & Schouten, M. (2007). *Exploring Humans. An Introduction to the Philosophy of the Social Sciences*. Amsterdam: Boom.

Dryden, J. (2010). Autonomy: Overview. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://www.iep.utm.edu/ autonomy/

Edmonds, D., & Warburton, N. (2012a). Adina Roskies on Neuroscience and Free Will. Retrieved on August 6, 2012, from http://philosophybites.com/2012/05/adina-roskies-on-neuroscience-and-free-will.html

Edmonds, D., & Warburton, N. (2012b). Neil Levy on Moral Responsibility and Consciousness. Retrieved on August 6, 2012, from http://philosophybites.com/2012/03/neil-levy-on-moral-responsibility-and-consciousness.html

Ekstrom, L. (2010). Volition and the Will. In T. O'Connor & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 99-107). Chichester: Wiley-Blackwell.

Eshleman, A. (2009). Moral Responsibility. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/moral-responsibility/

Fischer, J. M. (2007). Free Will and Moral Responsibility. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory* (pp. 321-354). New York: Oxford University Press.

Fischer, J. M. (2010). Responsibility and Autonomy. In T. O'Connor & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 309-316). Chichester: Wiley-Blackwell.

Garreau, J. (2006). *Radical Evolution. The Promise and Peril of Enhancing Our Minds, Our Bodies - and What It Means to Be Human*. New York: Broadway Books.

Haji, I. (2009). *Incompatibilism's Allure. Principal Arguments for Incompatibilism*. Plymouth: Broadview Press Ltd.

Haines, W. (2006). Consequentialism. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://www.iep.utm.edu/ conseque/

Hodgson, D. (2011). Quantum Physics, Consciousness, and Free Will. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 57-83). New York: Oxford University Press.

Hoefer, C. (2010). Causal Determinism. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/determinism-causal/

Hooker, B. (2008). Rule Consequentialism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/ consequentialism-rule/

Horgan, T. E., & Woodward, J. F. (1985). Folk Psychology Is Here to Stay. *Philosophical Review, 83*(2), 197-225.

Kane, R. (2011). Introduction: The Contours of Contemporary Free Will Debates. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 3-35). New York: Oxford University Press.

Knobe, J., & Nichols, S. (2011). Free Will and the Bounds of the Self. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 530-554). New York: Oxford University Press.

Latus, A. (2008). Moral Luck. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://www.iep.utm.edu/moralluc/

Levy, N. (2005). Libet's Impossible Demand. *Journal of Consciousness Studies, 12*(12), 67-76.

Levy, N., & McKenna, M. (2009). Recent Work on Free Will and Moral Responsibility. *Philosophy Compass, 4*(1), 96-133.

Lippert-Rasmussen, K. (2009). Justice and Bad Luck. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/ entries/justice-bad-luck/

Loewer, B. (1996). Freedom from Physics: Quantum Mechanics and Free Will. *Philosophical Topics, 24*(2), 91-112.

McKenna, M. (2009a). Compatibilism. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/compatibilism/

McKenna, M. (2009b). Compatibilism: State of the Art. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/compatibilism/supplement.html

McKenna, M. (2011). Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 175-198). New York: Oxford University Press.

McLeod, O. (2008). Desert. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/desert/

McNaughton, D. (1998). Consequentialism. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy* (Vol. 2). London: Routledge.

Meijsing, M. A. M. M. (2012). *Philosophy of Mind Week 7: Folk Psychology and Eliminative Materialism* [Sheets].

Mele, A. R. (2011). Free Will and Science. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 499-514). New York: Oxford University Press.

Moore, A. (2004). Hedonism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/hedonism/

Nahmias, E. (2010). Scientific Challenges to Free Will. In T. O'Connor & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 345-356). Chichester: Wiley-Blackwell.

Nahmias, E. (2011). Intuitions about Free Will, Determinism, and Bypassing. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 555-576). New York: Oxford University Press.

Nelkin, D. (2008). Moral Luck. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/moral-luck/

Nichols, S. (2008a). *Great Philosophical Debates: Free Will and Determinism (Guidebook)*. Chantilly: The Teaching Company.

Nichols, S. (2008b). *Great Philosophical Debates: Free Will and Determinism (Lectures)*. Chantilly: The Teaching Company.

O'Connor, T. (2010). Free Will. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/ freewill/

Pereboom, D. (2011). Free-Will Skepticism and Meaning in Life. In R. Kane (Ed.), *The Oxford*

Handbook of Free Will (pp. 407-424). New York: Oxford University Press.

Pink, T. (2010). Free Will and Determinism. In T. O'Connor & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 301-308). Chichester: Wiley-Blackwell.

Ramsey, W. (2007). Eliminative Materialism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/materialism-eliminative/

Russell, P. (2011). Moral Sense and the Foundations of Responsibility. In R. Kane (Ed.), *The Oxford Handbook of Free Will* (pp. 199-220). New York: Oxford University Press.

Searle, J.R. (2001). *Rationality in Action*. Cambridge, Mass.: MIT Press.

Sie, M., & Wouters, A. (2010). The BCN Challenge to Compatibilist Free Will and Personal Responsibility. *Neuroethics, 3*(2), 121-133.

Sinnott-Armstrong, W. (2011). Consequentialism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://plato.stanford.edu/ entries/consequentialism/

Sommers, T. (2009). More Work for Hard Incompatibilism. *Philosophy and Phenomenological Research, 79*(3), 511-521.

Stoljar, D. (2009). Physicalism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://plato.stanford.edu/entries/physicalism/

Strawson, G. (1998). Free Will. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy* (Vol. 3). London: Routledge

Timpe, K. (2006). Free Will. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://www.iep.utm.edu/freewill/

Vihvelin, K. (2011). Arguments for Incompatibilism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*. Retrieved on June 18, 2012, from http://plato.stanford.edu/ entries/incompatibilism-arguments/

Walter, H. (2011). Contributions of Neuroscience to the Free Will Debate: From Random Movement to Intelligible Action. In R. Kane (Ed.), *Oxford Handbook of Free Will. Second Edition* (pp. 515-529). New York: Oxford University Press.

Warfield, T. (2005). Compatibilism and Incompatibilism: Some Arguments. In D. W. Zimmerman & M. J. Loux (Eds.), *Oxford Handbook of Metaphysics* (pp. 613-630). New York: Oxford University Press.

Weijers, D. (2011). Hedonism. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://www.iep.utm.edu/hedonism/

Williams, G. (2006). Praise and Blame. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy.* Retrieved on August 6, 2012, from http://www.iep.utm.edu/ praise/