

One

RESPONSIBILITY, DETERMINISM, AND FREEDOM

His dream, though, is that some time in an unguarded moment—and this, it must be admitted, would require a night darker than any night has ever been yet—he will jump out of the fighting line and be promoted, on account of his experience in fighting, to the position of the umpire over his antagonists in their fight with each other.

“He” by Franz Kafka, cited by Hannah Arendt (Arendt, 1978, p. 202)

1. Introduction

This chapter introduces the three central concepts in the debate about responsibility—determinism, responsibility, and freedom—and explains what interpretation of them is relevant to this book. First, I argue that determinism is relevant in so far as it forces us to explicate and sharpen our ideas about the other two concepts, freedom and responsibility. I also argue that what should worry us is not that if determinism is true we are never able to act without interference or constraints. The real concern is the lack of originating powers of individual agents that the truth of determinism implies.

Second, I argue that questions about origination and ultimacy are connected with our status as accountable subjects. We need not fear that if determinism is true we can no longer be said to be agents (although we might worry about it). We can make a distinction between events and actions without using the concept of robust freedom and we can talk about human agency without using the concept of robust freedom. However, we sometimes feel responsible for our actions in a deep sense. We feel that besides being the cause of what we bring about, we are also truly to blame or praise for what we bring about. It is this deeper sense of responsibility that matters in our concerns about determinism. If determinism is true and we do not originate our actions (or that which causes our actions), we are merely the fortunate or unfortunate causes of the actions for which we are praised or blamed. So much moral luck should disturb us. Praise and blame are not the kind of sentiments that can or should be separated from considerations about individual desert.

Third, with regard to freedom, I argue that we have, as yet, no adequate analysis that dissolves the apparent incompatibility between freedom and determinism. We cannot explain away the necessity of freedom in so far as deep responsibility is concerned. All excellent attempts to show that determinism and freedom are compatible notwithstanding, determinism still appears to

threaten the kind of freedom necessary for deep responsibility. Therefore, we still need a story to tell about our daily practices of responsibility in so far as we believe that the praise or blame we bestow on individual agents is justified.

At the end of this chapter, I hope to have illuminated why we need to take a closer look at the view that I call “practical compatibilism.” Both prominent defenders of this view, Susan Wolf and R. Jay Wallace, acknowledge that previous philosophers have not solved the apparent incompatibility between determinism and deep responsibility. As a reaction they propose a new approach that appears promising in the light of our diagnosis of the problem, although, as I argue in subsequent chapters, it has its own problems.

Those who already feel at home in philosophical discussions about responsibility could start reading the third section of this chapter.

2. Determinism

Because many people involved in the discussion about our practices of responsibility claim not to understand precisely what the thesis of determinism is (Strawson, 1962, p. 59), we need some initial definition to start out with. Ted Honderich, who defends the intelligibility and the truth of determinism, defines it as the theory that—like everything else that happens—all our mental events and actions are the effects of things that happened in the past. They have to happen or are necessitated, and cannot be owed to origination. This absence of origination, as will become clear, is important. According to Honderich, it means that all our actions have a cause, which he understands as an event or condition within the causal circumstances. He understands the causal circumstances in turn as a set of conditions or events that necessitate or stand in nomic connection with something later, an effect. The causal circumstances make it impossible that any event other than the effect will occur (Honderich, 1993, pp. 11–18). The crucial point here is that determinism implies that whatever happens is determined by prior events, which are again determined by prior events, and so on.

According to Robert Kane any event is determined if conditions arise the joint occurrence of which is sufficient for the occurrence of the event (Kane, 1996, p. 8). If the joint occurrence of some conditions is sufficient for the occurrence of an event *a*, then given that these conditions obtain, *a* is inevitable. If determinism is true, all events, including our choices, decisions, and actions, are necessitated and no room exists for origination by a single individual. This means no clear and distinct activity of a single being exists of causing or initiating a mental event, a choice or decision, or an action.

Which conditions are to be regarded as the determining conditions depends upon the version of determinism in hand. For theological determinism, God’s omnipotence and what follows from His decrees are the determining conditions. For physical determinism, the determining conditions are the laws of nature together with the antecedent events. For psychological determinism,

the determining conditions are our motives and desires that in turn are understood to be the inevitable products of our birth and upbringing. What all versions equally lack is an “area of genuine agency,” as Thomas Nagel has put it (Nagel, 1979, p. 183). The view that everything is determined by a combination of laws linking events and prior events has no room for any activity that constitutes the agency of the individual involved in the chain of events.

Nowadays, the most cited version of determinism is physical determinism as defined by Peter van Inwagen. His definition is the conjunction of two theses:

- (a) for every instant of time, a proposition exists that expresses the state of the entire physical world at that moment, and
- (b) if x and y are propositions that express the state of the world at some instants, then x together with the laws of physics entails y .

In this definition it is taken for granted that the concept of a state does not logically entail its effects or causes, and that a change in things in the world necessarily implies a change in the state of the world. An advantage of this definition is that we need not define or know exactly what a law of physics is, as long as we agree that it cannot be rendered false by human beings. It then follows that if determinism is true, our acts are the consequences of the laws of nature and what happened in the remote past. But it is not up to us what went on before we were born, nor is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us (Van Inwagen, 1983, p. 16). This argument is called the “consequence argument.”

Not everyone is convinced of the validity of this argument, but possible objections to it play no significant role in the contemporary discussion about our practices of responsibility. Hence, with the exception of the objection of David Lewis that I sketch below (sec. 4.A.i), I will not discuss the argument. Determinism, no matter if and how we understand it exactly, sees the world in terms of causes and effects, causes and effects that are inevitable. That, as I show in the sections below, appears to be incompatible with our status as responsible human beings.

So why not reject determinism? As some have argued, physical determinism is in retreat anyway, for example in the physical sciences, where quantum theory has reintroduced indeterminism (Kane, 1996, p. 9). Unfortunately, that in itself is no solution, especially not if it is our status as responsible human beings that we are concerned with. The indeterminism of quantum physics concerns elementary particles that—just as the random swerves of the atoms that were once postulated by the Epicureans to account for the existence of free will—are a long way from providing the sort of freedom that we need for upholding human responsibility.

First, random swerves of elementary particles lose almost any significance in larger physical systems such as the human brain and body; indeterminism at a micro level does not surface in the behavior of our brains or

bodies. This means that an adapted version of determinism is still defensible. This version, called “near-determinism,” holds that while indeterminism exists at the micro level—the level of small particles—determinism still exists at the macro level, which includes neural events and everything with which we are ordinarily familiar (Honderich, 1993, p. 140). Second, even if the indeterminacies were occasionally to have large-scale effects on human behavior, this still would be of no help to the responsibility debate, for the effects caused by the indeterminacies would be unpredictable and uncontrollable, and unable to expand our responsibility. So this type of indeterminism provides no ready solution to the problem of responsibility. If we want to defend our practices of responsibility by arguing in favor of indeterminism we need better arguments (see Kane, 1996, Pt. 2; O’Connor, 1995; Searle, 2001, ch. 9).

Whereas most philosophers agree that indeterminism does not enhance our freedom and control, far less people believe that our freedom and control are threatened by determinism in the first place. These people are commonly called “compatibilists, their opponents “incompatibilists.” Since the thesis of this book is based upon the initial plausibility of such a threat—the initial plausibility of incompatibilism—let us take a closer look at those so-called compatibilist positions.

A. Control, Constraints, and Impediments

According to Daniel C. Dennett, our fear of a loss of control if physical determinism is true is unwarranted and completely misplaced. He argues that only the control that other agents exercise over us can diminish our freedom and lessen our responsibility. If someone makes us perform action *a*, we are not responsible for *a* (not as responsible as we would have been had we done *a* without anyone’s interference). Nature, though, Dennett argues, does not control us in that way. Nature does not make us do things as though we were puppets on a string. The reason we fear determinism, according to Dennett, is that our thinking about causation is infected with the feelings that accompany our idea of being controlled and interfered with by the malicious brain-manipulators and other ingenious scientists who populate the literature on responsibility. Once we understand that this is the wrong way to conceive of the manner in which nature determines us, we will see that our fear of determinism is mistaken (Dennett, 1984, pp. 61–66).

Dennett traces the origins of the topic of control and its relation to causation and determinism back to the early seventies (Dennett, 1984, p. 51). But this does not mean that the topic discussed under the heading of “control” is that new. The reason control has only recently come under discussion as such is probably that full-blown agent theories of freedom have only recently become influential in the discussion on responsibility. Like the notion of control, agent theories of freedom came to the fore in the early seventies with Harry G. Frankfurt’s (Frankfurt, 1971) and Gerald Dworkin’s (Dworkin, 1988)

elaborations on the hierarchical concept of a person. Unlike action theories of freedom, these theories treated freedom, free will, and autonomy primarily as characteristics of agents, which shifted the attention from the possibility of freely willed or autonomous actions in a deterministic universe to the possibility of autonomous agency or agents with free will in a deterministic universe. It is no wonder that Dennett's criticism—that incompatibilism suffers from a so-called animistic conception of causation—appeared in the debate on responsibility long before the notion of control came into fashion. Those who endorsed this kind of criticism are called "constraint compatibilists" (see, for example, Ayer, 1954, p. 22). Although constraint compatibilists focus on the freedom of individual actions rather than on the agent's control, the underlying idea was much the same: nature does not constrain us, and only constraint makes us unfree.

As I explain in section 4.B below, Dennett's compatibilism and the older version of it (constraint compatibilism) concentrate on a concept of responsibility that is called "superficial." The incompatibility between determinism and responsibility that I will concentrate on in this book does not confuse the difference between causes that impede the will and causes that do not. What troubles the type of incompatibilist that I discuss is that the mere absence of constraints and obvious impediments to our will does not guarantee that our actions are up to us; it does not guarantee that we are ultimately answerable for our actions. This concern is typical of what I call "ultimacy pessimism," because this position is pessimistic about the prospects of finding an adequate justification for our practices of responsibility (see Double, 1991, p. 365).

B. Ultimacy

Ultimacy pessimism holds that we are not responsible for those actions that can be traced back to an origin outside ourselves. It believes that if an action—let me stipulate that it is an unconstrained action—can be traced back completely to a combination of our upbringing, the culture in which we were brought up, and our biological constitution, we are not responsible for it because we are not responsible for things that lie outside our control. Ultimacy pessimism gives voice to the fear that if all our mental events (including our choices and decisions) and actions are effects of things that happened in the past, together with some determining conditions—as determinism states—the origin of our choices, decisions, and actions always lies outside ourselves.

On closer examination, as Nagel has put it, we are disturbingly subject to luck. We do not determine what kind of person we are and/or which inclinations, capacities, and temperament we have; nor do we determine the kind of problems and situations we face. We also do not determine how we are formed by antecedent circumstances or how our actions turn out (Nagel, 1979, p. 177). This is what is called "moral luck," and it is the existence of moral luck that makes the ultimacy pessimist so pessimistic. If even our unconstrained actions

are only effects, then what would make us responsible for them? As Roderick M. Chisholm writes: “If the flood of desire caused the weak-willed man to give in, just as the flood caused the poorly constructed dam to break, then just as the dam, the man is not responsible for the consequences” (Chisholm, 1964, p. 25).

Again, the simple rejection of determinism in favor of indeterminism cannot come to our rescue. Random swerves that somehow take place in our brains or bodies do not provide the kind of ultimacy we are looking for. What we want is to be the origin or the originator of the actions we are held responsible for. We want to matter as *agents* in the chain of events, not because something happens randomly inside our bodies, something that triggers other events to take place, but because we actively do something, something for which we are responsible.

Choosing between indeterminism and determinism does not solve the problem of responsibility and free will. Indeterminists have as much explaining to do as hard determinists have. The basic anxiety appears to be the difficulty in understanding our exact place as agents in a world of events (compare Chisholm, 1964, p. 24). Determinism primarily forces us to illuminate and sharpen our ideas about freedom, responsibility, and the relation between them. If everything is determined by prior events, then why do we treat human agency any differently than the agency of other beings with causal powers? And, if we do not believe that human agency is a special kind of agency, how do we defend our daily practices of responsibility in which we engage in a vocabulary of blame, praise, faults, desert, and merit, a vocabulary that we understand to be appropriate for humans but not for very young children, animals, and inanimate objects with causal powers? Those are the issues that are the subject of this book. So, what about this special kind of agency that figures in our daily practices of responsibility? How exactly are we to perceive of it?

3. Responsibility

One of the difficulties in discussing the relation between freedom and responsibility is that both concepts as we use them in our daily practices are ill-defined. Let us start with the concept of responsibility. We say, for instance, that the bank clerk who hands over the money to a robber in a cool and deliberate manner is not responsible for the bank’s loss of money: We feel that if someone has a gun held to the head, that person has no real options other than to comply with the demand made, and is not to blame for that action. However, if someone were to ask who was responsible for handing over the money to the robber, we would all point to the bank clerk, since the clerk was the person who handed over the money.

Also we could say that the bank clerk acted responsibly and made the right decision in handing over the money to the robber instead of trying to resist. If we say that the bank clerk acted responsibly, we mean that the clerk

acted appropriately, even in a praiseworthy manner, given the dangerous circumstances. Praiseworthiness and its opposite, blameworthiness, are in turn related to the concept of moral responsibility. Nevertheless, we would not say that the bank clerk was morally responsible for handing over the money to the robber, for in this case “morally responsible” means “blameworthy” and we do not think that the employee is blameworthy.

If, finally, we discover that the bank clerk is a programmed robot, we would still point to the robot as the one responsible for handing over the money, but we would skip all talk about moral responsibility, blame, and praiseworthiness. True and full moral responsibility is reserved for the restricted class of adult human beings.

We use “responsible” to refer to such apparently diverse things as “being the cause of x” (identifying the bank clerk as the person who handed over the money) and “being an appropriate subject of (moral) evaluation” (which the robot is not, but the human employee is). In the philosophical discussion on responsibility, the distinction between responsible in the “agent is the cause” sense and responsible in the “moral evaluation” sense is referred to as that between “superficial responsibility” and “deep responsibility” (see, for example, Wolf, 1990, pp. 40–41).

The example of the bank clerk also indicates another distinction, between “responsible” as a generic term referring to accountability, and “responsible” as a predicate that we use to evaluate the action as worthy of blame or praise. In the latter sense, “responsibly” and “morally responsible” can be used to refer to agents that should be praised for an act (they acted responsibly) or that should be morally blamed for an act (they are morally responsible). The first, generic sense of the term depends upon accountability conditions for its correct application, whereas the latter, evaluative sense depends upon blameworthiness conditions for its correct application (Wallace, 1994, p. 70).

In terms of the above distinctions, the topic responsibility as discussed in this book is that of deep accountability, the status of people as appropriate subjects of the reactive attitudes and emotions such as resentment, blame, indignation, and praise. I refer to these as “the reactive or moral sentiments”—the sentiments that were made influential through Peter F. Strawson’s discussion of responsibility (Strawson, 1962). This said, I return to the use of the terms “responsibility,” “deep responsibility,” and, sometimes, “moral responsibility” in a loose sense, as is common in the current debate. Whenever the distinction matters, I make clear which of the concepts is at stake.

A. Expectations and the Moral Sentiments

Wallace has developed a convincing reactive account of the relation between responsibility and the reactive attitudes and emotions. He introduces the notion of an “expectation” as the important mediating term (Wallace, 1994, chs. 2 and 3).

According to Wallace's account, holding someone responsible must be understood as holding someone to an expectation, which in turn should be understood as the susceptibility to a range of reactive attitudes and emotions with regard to that individual (although this susceptibility should not be understood as something that is conceptually prior to the holding responsible). People are susceptible to a range of reactive attitudes and emotions with regard to an individual if a wrong committed by this individual provokes:

- (1) a reaction with one of the moral sentiments, or
- (2) the judgment that a reaction with one of the moral sentiments would be appropriate, or
- (3) both (1) and (2) (Wallace, 1994, p. 21).

The inclusive disjunction is necessary because it is possible to hold someone responsible without experiencing any of the reactive attitudes or emotions.

Some people, for instance, are not very prone to moral sentiments. Others do not experience some of the reactive attitudes or sentiments—such as the negative ones of blame and moral indignation—with regard to a special group of people, for instance those they love or those they fear. Still others will only experience the moral sentiments with regard to the people they care about or love most. These possibilities do not necessarily imply that we do not hold the individuals concerned responsible for their behavior. It is not the actual reactive attitude or emotion alone that establishes that we hold someone responsible. However, when someone commits a wrong and we neither experience any reactive attitudes or emotions nor believe them to be appropriate, then there appears little reason to say that we hold this person responsible (Wallace, 1994, p. 23).

The general background condition of the inclusive disjunction is that the reactive attitudes and emotions are available as a reaction to the transgression of some expectations. We might imagine a society in which the reactive attitudes and emotions are not available. In that case the inclusive disjunction is of no use because nobody will in fact react with the moral sentiments and/or judge that it would be appropriate to do so. If we understand the availability of "a" reaction to the transgression of some expectations as crucial, instead of "the reactions that are available in our culture and which happen to be the moral sentiments," Wallace's reactive account is also compatible with, for instance, so-called shame cultures—which, it is claimed, do not know the sentiments of blame, praise, and indignation, but only that of shame (Wallace, 1994, pp. 36–37).

This compatibility with other cultures—especially ones in which the moral sentiments are not available—speaks strongly in favor of the reactive account. It enables us to embrace it even if we entertain an in itself healthy suspicion to the moral sentiments.

Another advantage of Wallace's reactive account is that it explains why we use the concept of responsibility in such diverse ways. This diversity is

partly due to the different domains of expectations: the legal domain, the moral domain, and the personal domain. Each of these is regulated by its own norms and values and, correspondingly, by its own normative expectations and demands. That we nevertheless use the term “responsible” in all domains can be explained by our shared susceptibility to some range of reactions—in our Western culture those of the reactive attitudes and emotions—to which the violation of a demand or expectation gives rise.

The range of attitudes and emotions that Wallace accepts as relevant to the stance of holding someone to an expectation is much narrower than that of Strawson. Strawson makes a distinction between three different kinds of reactive attitudes: those involving direct interpersonal relations (personal reactive attitudes like resentment and gratitude), the vicarious analogs of them that primarily concern someone’s behavior toward others (moral reactive attitudes like moral indignation), and the self-reactive attitudes like guilt and remorse (Strawson, 1962, pp. 62–64). Wallace excludes those reactive attitudes that relate more broadly to the interpersonal relationships, such as love, gratitude, hurt feelings, and shame, and restricts himself to those sentiments he calls the “moral sentiments,” the main ones being: moral blame, resentment, moral indignation, and moral praise (Wallace, 1994, pp. 25–29).

I accept this restriction, but without the addition of and the emphasis on moral. In my view, the main distinction (although it is not a robust one) is that between superficial and deep responsibility, not that between responsibility and moral responsibility. I explain this preference later on. First I explain the distinction between superficial and deep responsibility as it is normally used.

B. Superficial Responsibility

Superficial responsibility is the kind of responsibility that we share with all things that can be said to cause other things to happen (inanimate things, animals, artifacts, and human beings). When used to refer to something as a cause, “*p* is responsible for *a*” identifies the person or thing (*p*) to which something—a state of affairs, an event, or an action (*a*)—can be traced back. For example: The rain is responsible for the muddy streets. Correspondingly, we could believe that “*p* is responsible for *a*” means “*p* is the agent of *a*,” and that “*p* is the agent of *a*” means “*p* made the intentional bodily movements constituting *a*.” This would keep things very simple and straightforward because we often make bodily movements that cause things to happen. What could possibly be problematic about our daily practices of responsibility? The rain is responsible for the muddy streets and *p* is responsible for tracking the mud into the house. Unfortunately, this apparently unproblematic approach to our daily practices of responsibility fails to identify all that we want it to identify. Let me explain.

Our actions, in so far as they are intentional bodily movements, are not the only things for which we might be held responsible. We are also often held

responsible for omissions, things we did not do and which did not involve any bodily movements, for instance, forgetting our mother's birthday. Likewise, we are often held responsible for accidents or mistakes, bodily movements that are not intended—not under the description under which we are held responsible for them; for instance, stupidly hurting someone in our attempt to get into the train. Finally, we are often held responsible for careless and reckless behavior, behavior that need not necessarily involve intentional bodily movements; for instance, stumbling over a small child while running across a schoolyard. All of these are examples of actions for which we can be held responsible, but which cannot be translated into intentional bodily movements.

What appears to be crucial in our assessment of *p* as being responsible for *a* is not *p*'s (intentional) bodily movements that constitute *a* as such, but *p*'s bodily movements or lack thereof relative to a normative background. Rules, norms, and values determine what we should or should not do, and it is only relative to those that we are held responsible for actions and/or omissions. Even at the most superficial level, responsibility is connected with expectations concerning the appropriate behavior and actions in certain contexts. This is why not paying attention can be something for which we are held responsible, namely, if we believe that we should have paid attention in that particular situation. Likewise, making mistakes can be something for which we are held responsible, namely, if we believe that we should not make mistakes in that particular situation, and so on.

As soon as we go beyond our assessment of *p* as being “a mere cause of *a*,” we enter the normative domain. That is why the distinction between superficial and deep responsibility is not a very robust one. Both superficial and deep responsibility are connected with our ability to act according to norms, values, and expectations. The distinction nevertheless helps to focus our attention on a part of the discussion on responsibility: the part connected to autonomy. I come back to this in section 6 below and, more elaborately, in the following chapter.

For the moment, let me note that I will not discuss the question whether we are superficially responsible, whether we ever do things that we should not do and do so freely (where “freely” stands for: without being forced or interfered with by powers outside ourselves). It appears odd to seriously question this. It also appears odd to suppose that the distinction between superficial and deep responsibility could be the solution to a complicated philosophical discussion that has been going on for ages. Not that the activity of people who philosophize legitimizes itself, but the distinction is obvious. The incompatibilists and ultimacy pessimists central to this book reject the possibility of constraint compatibilism, discussed in the previous section.

The incompatibilism and ultimacy pessimism that I focus on agree with the constraint compatibilist's contention that a difference exists between being forced or constrained by external impediments, such as other human beings, and not being forced or constrained and, therefore, in some sense being free and responsible. What they do not believe is that this distinction alone suffices

to justify our daily practices of responsibility. What they are concerned with is whether our superficial responsibility for things ever reflects a deeper responsibility.

C. Deep and Moral Responsibility

With regard to responsibility, a class of distinctions exists that are related to the idea of autonomy. I call these “the autonomy-related distinctions.” Our use of the autonomy-related distinctions surfaces in the fact that we sometimes say of human beings that they are responsible for their actions in a way that animals, robots, very young children, mentally ill, or mentally disabled people are not. When we use “responsible” in this deeper sense, we claim that p should not have done a and that nothing prevented p from not doing a , and that p is, somehow, a fit subject of credit and discredit, that p deserves to be blamed or praised for a (Wolf, 1990, p. 41). Were we to discover, as in the case of the bank clerk, that p is a robot, or a child, or a mentally ill person, we would exempt p and no longer hold p deeply responsible for a .

The (in)compatibility of deep responsibility with determinism is often discussed under the heading of the (in)compatibility of moral responsibility with determinism. This is understandable, for moral responsibility—which is related to the *moral* expectations and demands that we hold people to—occupies a central and important place in our daily lives. It regulates large parts of our interpersonal relationships and affects our thinking about important social and political areas, such as social reform, law, and punishment. Deep responsibility need not necessarily be of a moral nature, though. As Wolf points out, we also judge the works of art of a mature artist differently than the finger paintings of a young child (without claiming anything about their beauty; Wolf, 1990, pp. 7–8). When judging and evaluating the work of someone we consider to be a deeply responsible artist, we take into account that it reflects artistic choices and preferences in a way that the work of, for instance, a young child does not. Both can make an ugly painting, but only the first will be held deeply responsible for doing so because the painting reflects the wrong artistic choices.

What appears crucial to the class of moral sentiments of blame, resentment, indignation, and praise is that each sentiment can be understood as a reaction to a perceived failing or failure of the agent, or, in the case of positive emotions and attitudes, as a reaction to a perceived accomplishment or achievement of the agent. And failures and achievements need not necessarily be of a moral nature: For instance, we can also blame someone for giving a bad piano performance.

Wallace rightly argues that condemning a piano performance for its emptiness or lack of intelligence is different from morally blaming someone for a wrong that is committed (Wallace, 1994, p. 54). Unfortunately, the example fails to establish what Wallace wants it to establish, which is the substantial

difference between moral and deep responsibility. Condemning a piano performance for its emptiness or lack of intelligence is not the correct analog of morally blaming someone. The analog of morally blaming someone for a wrong committed is blaming a pianist for a bad performance. If we blame someone for a bad performance, we add a force and quality that differs from condemning the performance, because we add that it was the pianist who failed to do what the pianist was able to do.

Blaming differs from merely describing a performance as “bad.” If people blame you for your bad piano performance, this is directly related to the expectations that they think you are capable of fulfilling. This is why blame—its special gravity and negativity notwithstanding—is sometimes preferred to the simple description of an action as “bad.” After all, if people do not blame you for the bad piano performance but merely comment on its poor quality, this might also reflect their low esteem of you or their lack of expectations regarding your abilities as a performer. Likewise, praise can be hurtful and insulting when it is a reaction to your exceeding someone’s expectations of you. Praising someone excessively for a mediocre performance is a very effective way of communicating our low esteem of the performer’s abilities to perform any better.

Wallace’s account is ambiguous with respect to the distinction between moral and deep responsibility. On the one hand, he explicitly states that moral responsibility can be distinguished by the “(. . .) *kind* of expectations it is bound up with” (Wallace, 1994, pp. 35–36). On the other hand, he emphasizes the special character, quality, and force of the *moral* responses (Wallace, 1994, p. 54). As may be clear, I abandon this second emphasis in favor of his first contention. Although moral responsibility may have a special importance to us, it is not only in the moral variant of deep responsibility that our responses to the actions and behavior of others acquire a distinctive quality and force. They acquire this distinctive quality and force because what they are reactions to is the perceived failing or succeeding of someone against the background of some domain of demands and expectations.

Let me recapitulate. *P* is blameworthy or praiseworthy for *a* if *p* is a responsible subject and if *p* failed to fulfill or succeeded in fulfilling some expectations by doing *a*. *P* is an appropriate subject of the moral sentiments of blame, praise, and resentment if *p* is a responsible subject. *P* is a responsible subject if *p* is able to fulfill several—yet to be described—expectations.

The connection between these three is made by Wallace’s definition of holding responsible: the susceptibility to some range of reactive emotions and attitudes (if these are available in the society we are concerned with) when expectations are breached. The susceptibility to this range of attitudes and emotions means that we react with these attitudes and emotions and/or find such a reaction appropriate. In the following chapter, I explain why Wallace concentrates on holding someone morally responsible, which he defines as: “[Holding] a person to moral expectations that one accepts” (Wallace, 1994, p. 51).

I am quite aware that these definitions, unrefined as they are, are contestable on several points. People can argue that the moral sentiments and emotions play a far more foundational role than is allowed by these definitions. See, for instance, Lars Hertzberg, who argues that a world without blame would require the abolition of everything involving rules and the teaching of rules. Such a world, he claims, would be devoid of all institutional behavior, including science and language. It would leave nothing that we would care to call a human society (Hertzberg, 1975, p. 501). From such a pragmatist perspective, the idea that the appropriateness of our moral sentiments should be related to *p*'s being a responsible subject, and *p*'s being a responsible subject to *p*'s ability to fulfill some (yet to be specified) expectations, is controversial. With regard to moral responsibility, pragmatists will argue that *p*'s being a responsible subject is exhausted by *p*'s being an appropriate subject of the moral sentiments, in whatever sense "appropriateness" is defined in our daily practices.

On the other hand, we might also wish to argue that the moral sentiments only have meaning if we are agents who possess free will. Paul Benson, for example, argues that the importance of free will remains unaltered if we concentrate on the question whether it would be rational to give up our understanding of others as free and responsible beings (Benson, 1990; his argument is a reaction to Wolf, 1981). We might also argue that the so-called appropriateness of our moral sentiments can only be defended on the basis of a full-blown ethical theory, such as utilitarianism. And so on.

Let me stress that the above definitions are not meant to conclude the debate about what the moral sentiments are or what exactly responsibility is. They are meant as a description of the battleground that is the subject of this book. The important questions are yet to be posed, for example the following: When exactly are we appropriate subjects of the moral sentiments? What does it mean to be able to fulfill expectations? Are we capable of fulfilling expectations? And so on. Let us proceed to the heart of the battleground: the conditions of responsibility.

4. Freedom

We hold people responsible if we hold them to some expectations. We hold them to some expectations if we are susceptible to the moral sentiments with regard to them. Our susceptibility to the moral sentiments is justified if they are capable of fulfilling the expectations we have of them. In order to be capable of fulfilling these expectations, we need to:

- (1) have the abilities that are relevant to fulfill these expectations, and
- (2) be in the circumstances that allow us to fulfill these expectations.

It is a widespread conviction that people who breached our expectations are not

responsible for this if they could not have done otherwise. This intuition is philosophically translated into the so-called principle of alternative possibilities (hereafter PAP): A person p is only responsible for an action a if that person could have done something other than a .

How should we understand this counterfactual freedom? Take again my example of the bank clerk from the previous section. Let us for convenience suppose the clerk is a woman called Mary. Mary is held (superficially) responsible for handing over the money, because she could also not have handed over the money. But she was not (deeply) responsible, because she could not have done otherwise. She had no reasonable alternative to handing over the money. It is not completely clear what we mean when we say that people “could or could not have done otherwise than they did.”

A. The Conditional Analysis

One influential proposal is that we should analyze “ p could have done otherwise” as “ p would have done otherwise if p had willed, chosen or decided to do so.” This captures the first superficial sense in which our bank clerk could have done otherwise. Nothing prevented Mary from handing over the money. She handed over the money in a cool and deliberate way, so had she decided otherwise, she could have done something else than hand over the money. This analysis of our ability to do otherwise is known as “the conditional analysis” and is commonly ascribed to George Edward Moore (Moore, 1911). According to Wallace, though, it can be traced back in the empiricist tradition as far as Thomas Hobbes’s *Of Liberty and Necessity* (Wallace, 1994, p. 193).

A great advantage of this analysis is that it is compatible with determinism. Mary’s unexercised freedom to do otherwise is not in contradiction with her having been determined to do what she did, because had she chosen, willed, or decided to do otherwise, the causal circumstances in the past would have been different. It is still true that for every event conditions exist the joint occurrence of which is sufficient for the occurrence of that event.

A common objection to this analysis is that it fails to do justice to issues concerning our freedom of will. After all, Mary could not have done anything other than hand over the money to the robber in the second deep sense of the above example (see Watson, 1987a, p. 160). She could not have decided, willed, or chosen not to hand over the money, because she had a gun pointed at her head. If she could not have decided, willed or chosen otherwise, then she could not have done anything other than hand over the money, whereas it is still true that she would have done otherwise if she had chosen, willed, or decided to do otherwise. The conditional analysis fails for it is possible that the *analysans* “ p would have done anything other than a if p had chosen, willed, or decided to” is true, and the *analysandum* “ p is able to do a ” is not (see Chisholm, 1964 and Lehrer, 1968; for objections see Wolf, 1990, p. 99).

We could doubt whether the second sense in which the bank clerk is

unable not to hand over the money—that is, the sense in which she has no reasonable alternative—is a real inability. The fact that people are unable to choose, decide, or will something does not mean that they are unable to do the thing that they are unable to choose, will, or decide to do. If someone is offered fifty dollars to tell the time, the added incentive will render it relatively impossible not to tell the time. But this inability is not a real inability not to tell the time. The offer does not render us incapable of not telling the time. (The example is Wolf's, 1990, p. 98.) Not all senses of “not being able to” constitute a robust inability and not all, so it appears, equally undermine our freedom.

This fact is partly obscured when we use examples in which extreme offers or threats figure, because offers you cannot refuse and threats are manipulative measures that are by definition understood as freedom-undermining: if they are not freedom-undermining, they are not manipulative. I come back to this elaborately in the next chapter (sec. 3.B). To prevent this prior assumption from obscuring matters, we should imagine someone choosing or deciding to do something that renders them unable to do anything else. If people hold their breath, for instance, it becomes relatively impossible for them to do anything else. This is not a real inability that affects people's freedom in any relevant sense (Aune, 1967, p. 38). Only causes that impede the will are freedom-undermining. What appears to matter most is whether people do something of their own free will; that is, whether they would have done otherwise if they had chosen, decided, or willed to do so.

What happens if our bank clerk is panic-stricken? Not panicking physically—for then her handing over the money would not be an action—but she could be in such a shock that she hands over the money psychologically blinded and shaken by fear. If this is the case, she could not have chosen, willed, or decided to do otherwise even if she would have done otherwise had she chosen, willed, or decided to do so (compare Wolf, 1990, p. 99). This would again falsify the conditional analysis, for it proves that the *analysans* can be true (she would have done otherwise if) while at the same time the *analysandum* (she could have done otherwise) is not true. A person's counterfactual freedom to do anything other than *a* in the conditional analysis sense cannot be what makes that person responsible for *a*, for cases occur in which someone is in this sense free, but not responsible.

This last objection to the conditional analysis is widely accepted, although one final reply to the objection is worth mentioning. The validity of the above objection depends on how we analyze the activities of deciding, willing, and choosing. We could argue that these activities are not voluntary activities, that they are themselves determined by a variety of factors. We choose, want, or decide to go to the movies because, for example, we desire to relax a bit and believe that going to the movies will help us relax; relaxation is what we need most at this moment; we can afford the admission; we will not miss a terribly important telephone call; and so on. Our choices, volitions, and decisions are determined by desires, reasons, and beliefs; they are not made on no basis at all.

The conditional analysis could perhaps be improved and successfully defended if in the *analysans* we substitute “if she had chosen, willed, or decided to” with “if she had sufficient reason to.” The ability to do otherwise could then be analyzed as: would have done otherwise had someone had sufficient reason to do otherwise. This analysis would help us handle the difference between our cool and deliberate bank clerk and our panic-stricken one. If the cool and deliberate Mary had had sufficient reason not to hand over the money, she would not have handed over the money. The panic-stricken Mary, on the other hand, would still have handed over the money, even if she knew that handing over the money would mean getting herself killed (for instance, because she had seen the face of the robber, and the robber now has no option but to shoot her). In this rendering of the conditional analysis, cool Mary’s action is free, whereas panic-stricken Mary’s action is not.

Unfortunately, having sufficient reason to do otherwise cannot account for all cases in which we base our judgment that someone is responsible for an action on the presence of counterfactual freedom. It cannot account, for instance, for the category of weak-willed actions in which people have sufficient reason not to do something and recognize these reasons as sufficient reasons, but go ahead and do it anyway (see Fischer, 1986, p. 18). If I pour myself a glass of wine and drink it, whereas I have every reason to call it a night and am aware of these reasons—I even mentioned them while pouring the wine—there appears to be no reason to release me from responsibility. In these cases I appear to be free enough to be held responsible, although it is not true that I would have done otherwise had I had sufficient reason to do so. I had sufficient reason to do otherwise! I come back to weakness-of-will explanations in the last two chapters of this book (ch. 3, sec. 4.C and ch. 4, sec. 5.A).

If we do not believe in the existence of something like weak-willed action, still another class of actions exists in which our counterfactual freedom cannot be analyzed conditionally. These are our actions that failed to succeed without apparent reason: For instance, if we failed to keep our temper; or to score in a soccer game, despite the fact that all circumstances were ideal. No matter which conditional circumstances you believe in—be it choices and decisions or sufficient reasons—what you blame yourself for is failing *simpliciter*. You do not mean that you could have scored or kept your temper if the causal circumstances had been different, but that you could have scored or kept your temper in exactly the same circumstances (Austin, 1956, p. 218).

Various points raised in this section will reappear in the third chapter, where I explain why I believe this intuition with regard to the necessity of alternative possibilities to be such a strong and persistent one. For the moment, let me proceed by explaining why the debate about the conditional analysis ended with van Inwagen’s consequence argument.

i. The Consequence Argument

The discussion on the conditional analysis was concluded, more or less, with Van Inwagen's consequence argument. In the consequence argument, the thesis of determinism is expressed in a very simple and straightforward way with a minimum of assumptions, while the lack of an area of genuine agency is immediately clear. This renders the controversy surrounding the conditional analysis superfluous, because nothing in the consequence argument depends upon the interpretation of our ability to do otherwise—not at first sight. Lewis, however, criticizes Van Inwagen's consequence argument and argues that the real consequence of it does depend on the interpretation of our ability to do otherwise (Lewis, 1981). Remember the consequence argument stated in the second section:

If determinism is true, our acts are the consequences of the laws of nature and what happened in the remote past. But it is not up to us what went on before we were born; nor is it up to us what the laws of nature are. Therefore, the consequences of these things (including our current acts) are not up to us (Van Inwagen, 1983, p. 16).

According to Lewis, the worrying consequence of this formula is that if determinism is true, we cannot be responsible for anything we do, because we are unable to change the laws of nature or the facts of the past. He argues, that the actual consequence Van Inwagen's argument establishes is not this strong one that constitutes incompatibilism, but a much weaker one. This weaker consequence is that if determinism is true, we are only responsible if we are sometimes able to render a proposition that expresses a law of physics false or to render a proposition that expresses the state of the world at some point in time false. Although it is obvious that we cannot render a law of physics false or change the past (the strong consequence), it is not so obvious that we cannot render a proposition false that expresses either a law of physics or the state of the world at some point in time. The intelligibility of the latter, weaker consequence depends again upon our interpretation of the ability to do otherwise than we in fact did.

Nonetheless, Lewis's objection has not led to a reopening of the discussion about the conditional analysis. The consensus appears to be that the conditional analysis has proven to be unsuccessful. It fails to identify the freedom that is necessary for responsibility, and it equally fails to identify the freedom that is sufficient for responsibility. It fails to do the latter because cases exist in which we possess the conditionally analyzed freedom to do anything other than *a*, but are not responsible for *a*. This is the case, for instance, if we suffer from so-called unfreedom of the will. It fails to do the former because cases occur in which we lack the conditionally analyzed freedom to do anything other than *a*, but are still responsible for *a*. This is the case, for instance, if we render ourselves unable to perform some actions.

Since I do not believe in the second influential strategy for avoiding incompatibilism that I am about to discuss, I believe that the discussion about the exact interpretation of our ability to do otherwise is still very important. I will not, however, discuss this issue in much detail, as in this book my interests are restricted to the question of freedom only in so far as it is related to the justification of our daily practices of responsibility. And, as I argue later on in this book, to justify our practices of responsibility we need not solve the issue about the exact interpretation of our ability to do otherwise. What I hope to have established in this section, though, is the lack of easy solutions to the problem of justifying our daily practices of responsibility via a common sense compatibility of what we mean by “ p is responsible for a because p could have done something other than a ” and “ p could not have done anything other than a if determinism is true.” This brings us to the second influential strategy to avoid incompatibilism.

B. The Principle of Alternative Possibilities

Remember PAP as stated at the beginning of this section: A person p is only responsible for an action a if p could have done something other than a . The first to launch a straightforward attack on this principle was Harry G. Frankfurt (Frankfurt, 1969). Frankfurt argues that it is the efficient cause that is relevant to the moral assessment of a person, and not the availability or lack of alternative courses of action (see also Dennett, 1984, ch. 6). If p does a because p wants to do a , a lack of alternatives does not lessen p 's responsibility for what p does; hence, PAP is false. Frankfurt illustrates his point with two very powerful thought experiments, both involving a so-called counterfactual intervener: someone (or something) that would intervene to ensure that person p does action a , but who (or which) does not intervene because p does a without intervention. Following Van Inwagen, I call these examples “Frankfurt counterexamples” (Van Inwagen, 1978, p. 202), even though Frankfurt himself credits Robert Nozick as the originator of the counterexamples for which he became famous (Frankfurt, 1969, p. 6).

Suppose Jones is kept hostage in her house by a gangster and forced to ignore the cries for help that she hears outside. The threats of the nasty looking gangster make an impression on her and would force her to obey had she not already decided to stay in the house no matter what happens, because she desperately needs to finish the work brought home. If this is the case, two things are true: (1) Jones does not have a real alternative to staying in the house, because the threats make her stay inside even if she had not decided to finish the work; and (2) Jones is morally responsible for staying in the house since the threats do not exert any force on her whatsoever. She stays in the house because she wants to finish the work brought home.

We could object to this interpretation of Jones's so-called inability to do otherwise. Although Jones would have succumbed to the threats had she not

already decided to do what she was about to be coerced to do, so the objection would run, it is not true that she could not do otherwise. Jones is not, for instance, paralyzed. Frankfurt anticipates this objection by developing a second example around a jack-of-all-trades called Black. Black, by stipulation, is able to take sufficient measures to pre-empt all alternative possibilities of Jones's doing *a*, but does not need to because Jones does *a* without Black's intervention. Black's non-intervening presence secures the absence of alternative possibilities without at the same time undermining Jones's responsibility, because Jones does all the nice or nasty things Black wants her to do without Black's actual intervention.

The point of this example is compelling, for it is unclear how Black's simple looming in the background can affect Jones's responsibility for what she does freely and willingly. No actual difference exists between Jones's doing *a* in a world without Black's presence and her doing *a* in a world with Black's presence. The only difference is a counterfactual one: If Jones had tried to do something other than *a*, she would have found herself unable to do so.

Also convincing is Frankfurt's explanation of the confusion that lies at the origin of PAP. According to Frankfurt, the attraction of PAP derives from a too literal and mistaken interpretation of the class of excuses that express our inability to do one thing or another. Frankfurt points out that people who claim that they could not have done otherwise are only excused if their inability is a complete explanation of their action, that is, if it explains why they did not do otherwise. The claim that they could not have done otherwise means that they would have done otherwise if they had been able to do so. If we discovered that the lack of alternative courses of action had in fact no influence whatsoever on someone's behavior, we would no longer accept the inability as an excuse. Frankfurt concludes that PAP should be replaced by the following principle: You are not morally responsible for what you have done if you did it "only because" you could not have done otherwise. And this only-because principle, he claims, is compatible with determinism. I deal elaborately with this only-because principle in the next chapter (sec. 3.A).

Before we continue our discussion of PAP let me note that although the above example applies only to blameworthy actions, I do not believe as, for instance, Wolf does, any fundamental asymmetry exists between blameworthy and praiseworthy actions in this respect (Wolf, 1980; Wolf, 1990, ch. 4). I believe that our intuitions with regard to praiseworthy actions are less clear and explicit, probably because we have less reason to be cautious and mindful when praising others than we have to be with the negative normative sentiments. Praising someone does little harm even if the person in question does not deserve it, whereas blame is harmful. The choice to restrict our attention to blameworthy actions should be understood to reflect a concern with harming people even if they do not deserve it.

Frankfurt's argument against PAP initiated an elaborate and complex controversy that is still going on. I will not give an overview of the highly

complicated arguments and issues that have been raised in connection with it, but restrict myself to some preliminary explorations of the connection of PAP with the notions of ultimacy and autonomy, both of which are important for the conceptual argument in favor of PAP that I develop in the third chapter, and the ultimacy pessimism that I discuss in the fourth chapter. For this purpose, I use John Martin Fischer's excellent overview of an important line of argument used by those who disagree with Frankfurt counterexamples (Fischer, 1994); I also use Robert Kane's ultimacy principle (Kane, 1996).

i. Ultimacy

Kane argues that incompatibilists should accept Frankfurt counterexamples and admit that PAP is not in every instance necessary in order to render someone responsible for *a*. Consequently, incompatibilists should argue—as Kane himself does—for another principle that mentions alternative possibilities as a necessary condition for responsibility. He calls this principle the ultimacy principle (UP).

UP states that the availability of alternative possibilities is necessary at some stage in a person's life history in order for this person to be responsible for *a* at *t*₁. A person can be responsible for *a* at *t*₁ even without having any alternative possibilities, but only "(. . .) if he was responsible for making himself the sort of person that he then [at *t*₁] was by virtue of other choices or actions in his life history with respect to which he could have done otherwise" (Kane, 1996, p. 42). If determinism is incompatible with this principle—as Kane, being an incompatibilist, believes—then the justification issue is back on the agenda.

I prefer to put Kane's UP differently, but let me first explain why I think that Kane's proposal to substitute the ultimacy principle for PAP is correct. We can argue that *p* is responsible for *a* even if at the moment of *a*, or the moments preceding *a*, no alternative possibilities were present (against PAP), as long as somewhere in the chain of events leading to *a* at *t*₁, *p* could have done something other than what *p* in fact did (in favor of UP). If the latter is the case, we can translate *p*'s freedom to do otherwise at some instant *t*₁ (provided that it is relevant to the chain of events leading to *a*) into *p*'s freedom to prevent *a* from happening and, consequently, into *p*'s responsibility for *a*.

Suppose, for instance, that someone who is unable to swim impersonates a lifeguard. The example is Walter Glannon's (Glannon, 1995, p. 269). Confronted with a swimmer who gets into trouble, this lifeguard is unable to come to the swimmer's rescue. In this case, given Kane's ultimacy principle, this person can be said to be responsible for the drowning of the swimmer if somewhere in the history leading up to the drowning, the lifeguard could have done something, anything at all, to rescue the swimmer. This corresponds with my intuitions about the case. If, for instance, the impersonation was just a joke or an amusement and the impersonator knew the risks involved but still decided to

go ahead with the joke, we will blame that person for what happened regardless of the lack of alternatives at the moment of the action. If, on the other hand, no alternative possibilities were present at all in the chain of events that eventually led to the drowning, we would not hold that person responsible. This could be the case if, for instance, the inclination to impersonate was irresistible as the result of, say, a very peculiar and rare neurosis that drives people to imitate figures of authority (such as, in this case, a lifeguard).

Practical compatibilism focuses on the autonomy-related distinctions, in this case on the impersonator's status as an appropriate subject of the moral sentiments. Frankfurt counterexamples, in this respect, beg the question. The counterexamples assume that the person who figures in the example is an autonomous being. Let us explain this briefly with the help of Fischer's discussion of what he has called the "flicker of freedom strategy," a strategy employed by those who disagree with the Frankfurt counterexamples (Fischer, 1994, ch. 7).

ii. Autonomy

The flicker of freedom theorist argues that Frankfurt counterexamples are invalid because they presuppose some flicker of freedom. They would argue that my conclusion that Jones, in the example above, is responsible despite the presence of the counterfactual intervener, is provoked by this flicker of freedom. Hence, the conclusion that we are responsible regardless of the absence of alternative possibilities is not warranted, because even the tiniest flicker of freedom constitutes the presence of alternative possibilities.

Suppose that Black installs a mechanism in Jones's brain that will enable him to intervene in case Jones chooses or is inclined to choose to do something other than *a*. Also suppose, that Jones does all the nasty things Black wants her to do without Black interfering. According to the flicker of freedom theorist, Jones is only responsible for *a* if she—at the very minimum—has the freedom to initiate a different decision, that is, to choose or be inclined to choose to do something other than *a*. This initiative would perhaps be barely visible because it would immediately call for an intervention by Black, but it is, nevertheless, a flicker of freedom.

Such a flicker of freedom, this counter-objection would run, must always be present in the alternative sequence (the course of action that Jones does not take), no matter how sophisticated the example is. Hence, our intuition that Jones is responsible because she chose to do *a* is still accounted for in terms of the presence of an alternative sequence, albeit not a very robust one.

Another counter-objection of the flicker of freedom theorist is that Jones must have the power to bring about a different event. This objection depends on an essentialist principle of event individuation in which events are distinguished from each other on the grounds of their causal history. If *a* has Black intervening in its causal history, it is a different event from an *a* that does not

have Black intervening in its causal history.

We could object to such an essentialist rendering of event individuation, as Frankfurt himself does. But such an objection would be highly controversial and, with regard to the problem at hand, intuitively implausible. With regard to someone's responsibility for a particular event, the causal history of this event appears relevant and the difference between Black's intervention and non-intervention is obvious. And, supposing we want to hold Jones responsible for bringing about a state of affairs—instead of holding her responsible for a particular event to which the causal history matters—we could wonder, as Peter van Inwagen has rightly argued, whether we should still hold her responsible for that state of affairs if we know it would have obtained no matter what (Van Inwagen, 1978).

According to Fischer both counter-arguments aimed at defending PAP employ a notion of alternatives that is too thin to account for our ascriptions of responsibility (Fischer discusses four, but only these two are relevant for our purposes: Fischer, 1994, pp. 138–139). A flicker of freedom that only exists in the alternative sequence cannot possibly render us responsible for what we did in the actual course of action, let alone in an alternative course of action in which we would act unfreely. In order to defend PAP, we need more robust alternative courses of action (Fischer, 1994, p. 142).

This argument transfers the burden of proof incorrectly. To see this, we should keep the sequence of the argument in mind. We start with an important and widespread intuition that our practices of moral responsibility cannot be understood without the availability of alternative possibilities. Frankfurt develops a powerful example in which doubt is cast on this intuition. Consequently, some people—namely, flicker of freedom theorists—want to verify whether this example directs our intuitions in the right way, regardless of the absence of alternative possibilities. What they want to check is whether the Frankfurt counterexamples justify the conclusion that occasions arise on which we believe someone to be responsible despite the absence of alternative possibilities. If this is not the case, the force of the Frankfurt counterexamples would disappear. If we still intuitively assumed some freedom (even the tiniest flicker) to be present in a Frankfurt counterexample, then the example would presuppose the freedom the superfluity of which it set out to establish. This does not necessarily mean that the flicker of freedom that is presupposed is itself enough to ground our ascriptions of moral responsibility.

Suppose someone argues that dollars (freedom) are not needed to buy bread (responsibility), and someone else objects that this can only be established if examples involving the purchase of bread (responsibility) do not presuppose the presence of dimes (a flicker of freedom), then, in this dispute, the argument that dimes (a flicker of freedom) do not buy us bread (responsibility) is the wrong kind of argument. Even if dimes do not buy us bread, we need the dimes to get, via the dollars, to the bread. More importantly, we cannot argue for the redundancy of dollars with an example that accepts the need for dimes.

Like the flicker of freedom theorist, both the practical compatibilist and

the ultimacy pessimist object to the question-begging element of the Frankfurt counterexamples with regard to deep responsibility. If we restrict our attention to superficial responsibility, it is right to say that only the efficient causes matter, not the available or absent alternatives. We can admit that we are superficially responsible for *a* if we intentionally and voluntarily bring about *a*. When we want to discuss whether someone is deeply responsible for *a* if determinism is true, Frankfurt counterexamples are question-begging. They beg the question whether deeply responsible beings exist even if determinism is true, if they already assume that such beings exist even if determinism is true. But this is what they assume if the flicker of freedom that is found in the examples by the flicker of freedom theorist is a token of the assumption that we are dealing with normal human beings like ourselves.

The assumption that people are normal human beings like ourselves could equal the assumption that they are free human beings, who are normally able to act in other ways than they did and who are deeply responsible for their actions and behavior. The prior assumption that people are free and responsible, as may be clear, would render inert the fact that we believe them to be free and responsible, regardless of the presence of a counterfactual intervener. For, in this case, this assumption no longer establishes the redundancy of alternative possibilities for our practices of responsibility in general.

If we know that people are normal responsible human beings, we know that they are responsible for an action *a* if that action is correctly related to them: If they do *a* because they want to do *a* (and not because it is caused by external interference or constraint). The distinction between acting on your own and acting in spite of yourself under the influence of visible external impediments—as interference and constraints most of the time are—is clear and distinct (I come back to this in the next chapter, sec. 3.B.i). But what about the less visible and discernible ultimacy-undermining factors? Why should covert non-constraining control—manipulation that we are not aware of—undermine our responsibility for the resulting behavior (as we normally assume), whereas our psychological make-up and genetic inheritance (a general condition for all of us if determinism is true) do not?

What are the general conditions that allow us to distinguish between a responsible and a non-responsible human being? What is the difference between someone who is manipulated or physically determined to want to do *a* and to do *a*, and someone who is not? These are the questions a skeptic will continue to ask. Even Frankfurt himself remarks that people who do *a* only because they could not do otherwise are not responsible for *a*, even if they wanted to do *a* (Frankfurt, 1969, p. 10). But how can we ever establish that people do not act for the mere reason that they could not do otherwise, if the fact that we want to do something (or even want to do something badly) is not enough? I discuss this point extensively in the following chapters. For the moment, let me introduce the skeptical challenge that motivates the objections to the Frankfurt counterexamples and related objections to, for instance, constraint compatibilism, which was discussed in section 2.

5. The Skeptical Challenge

Like constraint compatibilism, views such as Frankfurt's link our responsibility to a distinction we all experience at one time or another: that between doing something because we are forced to do so and doing something because we want to do so. This distinction does not establish that when we believe we do *a* because we want to do it, we do it for that reason (and not because we could not do otherwise). Also, even if we could establish that we do *a* because we want to do *a*, we cannot establish that we could have done something other than *a*: Perhaps we could not have wished anything else, given our personal history and the laws of nature.

We understand ourselves as rational, responsible, and free creatures and it proves extremely difficult, if not impossible, to see ourselves otherwise. We "have to act under the idea of freedom," as Immanuel Kant is often cited as saying. Theoretically we might fancy the idea that we are unfree, but we cannot make any choice without regarding ourselves as free (Korsgaard, 1996, pp. 162–163). Although this idea is a stronghold of philosophers who want to defend the compatibility of freedom and determinism, it may equally be used in an argument against it. If it is impossible to act otherwise than under the idea of freedom, the experience of freedom offers no proof of freedom's existence.

Many desires, choices, or beliefs exist that we experience as alien and that are best described as "external" to ourselves. Take for instance the experiences of compulsive or phobic neurotics and unwilling addicts. But these are not the experiences that matter to the skeptic. The experiences (first person) or observations (second or third person) that sustain the skeptical challenge are those in which we feel free and responsible, even though we are not. If it is possible to feel free and responsible without being so, and if determinism excludes our being free, we may end up concluding that determinism does undermine the legitimacy of our daily practices of moral responsibility regardless of our so-called experience of freedom.

Unfortunately, we cannot rule out the possibility that some of the actions we experience as being motivated by the content of some mental state are, in fact, caused by something else, something beyond our power. Striking examples of such illusions of freedom and responsibility are those in which a so-called covert non-constraining control is exercised over us, by way of behavioral engineering and manipulation. Aldous Huxley's *Brave New World* and Burrhus Frederic Skinner's *Walden II* are fictional illustrations of this phenomenon. As Kane observes, these examples—or rather the unease they provoke—reveal the importance we attach to something like ultimacy (Kane, 1996, p. 65). Although people may be living in our *Brave New World* or *Walden II* who never do anything they would not do in a different so-called free world, we do believe (or, some of us do) that they lack something that is worth wanting. In the end, we might conclude that we are mistaken about this—that, on second thought, nothing worth wanting exists that is missing from the *Walden* and *Brave New Worlds*—but it is significant that the belief

itself is not incoherent.

We also cannot dismiss the fact that we can gain knowledge of the real causes of people's (or our own) choices, values, and reasons, by investigating their background or biological constitution. Some choices, values, and reasons are caused by, for example, a delusional psychosis, extreme personal stress, or some other psychological disorder due to, for instance, a very strict upbringing. It is quite possible that there was a time in which we did not recognize these causes and instead believed the choices, values, and reasons were made solely by the agent.

Where would this line of questioning end? Does it have limits in nature itself? We do not know whether, once we have scrutinized all the possible causes of our behavior and actions, the story of humankind will leave enough room for deeply responsible individuals who can be wholeheartedly praised or blamed for their actions and behavior (I examine this question in ch. 3). This brings me to the theory that I criticize, refine, and defend in this book, and which I refer to as "practical compatibilism" or, in its adapted version, "practical semi-compatibilism."

This version of practical compatibilism is not the same as what Christine M. Korsgaard has likewise called "practical compatibilism," a position she ascribes to Kant and which she calls "practical" because Kant was not a theoretical compatibilist: He did not believe that free will and determinism could be reconciled from a single point of view (Korsgaard, 1996, p. 209). What it refers to is the inevitability of our practices of responsibility, regardless of determinism's truth. I elaborate on this in the following chapter. Also, the label "practical semi-compatibilism" should not be confused with Fischer's position called "semi-compatibilism." Fischer holds moral responsibility to be compatible with determinism while at the same time arguing that determinism rules out all alternative courses of action and that we can never act otherwise than we do (Fischer, 1994). As will become clear I do not argue for this. What distinguishes practical compatibilism from more traditional kinds of compatibilism is this: It tries to evade the skeptical challenge by treating responsibility as a primitive, basic concept—primitive or basic not in the sense that it cannot be analyzed any further, but in the sense that no matter what theoretical insights we gain with regard to determinism and (in)compatibilism, there will always be people to whom the predicate "responsible" applies.

I have a proposal about exactly how this might be understood in the fourth chapter, but for the moment let it suffice to say that like practical compatibilists it argues that the questions that can be raised concerning our practices of responsibility are, as Wolf has put it, not about the conditions that validate them but about the conditions that regulate them. If the conditions we explicate prove to be either incoherent or impossible to satisfy, we should abandon or revise those conditions, and not our practices (Wolf, 1990, pp. 16–17). Consequently, it is not the existence of (deeply) responsible human beings that is at stake, but the fairness of the conditions that we use to establish (deep) responsibility. As Wallace puts it: "(. . .) the [compatibilist-incompatibilist]

debate should be seen as a normative debate, about the conditions that render it appropriate to hold a person morally responsible” (Wallace, 1994, p. 85).

With this understanding of the discussion, as I argue next, practical compatibilism appears to change the order of arguments. We do not examine the metaphysical assumptions of our daily practices of responsibility, nor do we discuss the (in)compatibility of determinism with them or its relevance to them, but we discuss the conditions of responsibility, in the assumption that they are, at most, unfair if determinism is true. Before going into the exact differences and similarities between the traditional and practical (in)compatibilist positions, I give an overview of the positions that are possible with regard to determinism, responsibility, and freedom.

6. (In)compatibilism

In the eyes of someone who concentrates on the tension between moral responsibility and determinism, the debate is concerned with two (in)compatibilities: that between responsibility and determinism—hereafter I call the resulting position “(in)compatibilism (R)”—and that between freedom and determinism—hereafter I call the resulting position “(in)compatibilism (F).”

(A) In order to argue for incompatibilism (R), the following options are available.

(1) Incompatibilism (F) is true and the freedom that is denied by determinism is the same freedom that is necessary for responsibility (which is most of the time taken for granted). The most influential representative of this position is Van Inwagen (Van Inwagen, 1983).

Those who believe that they have established incompatibilism (F) can draw the following two conclusions.

(a) Determinism is true; we are not responsible beings. This position is called “hard determinism” and few people defend it as such. Nevertheless, some defend a version of it, for example, Honderich (Honderich, 1993) and, more recently, Derk Pereboom (Pereboom, 2001).

(b) Determinism is not true; we are free and responsible beings. Those who draw this conclusion are called libertarians. A recent defender of this position is Kane (Kane, 1996). Although the position had been out of fashion for quite a few years, it now appears to be back on the philosophical agenda (Watson, 2003, chs. 5, 9, 15). A big issue for libertarians is to explain how indeterminism can make us responsible for our actions.

(2) A second position is defended by, among others, Chisholm, who argues that both determinism and indeterminism are incompatible with responsibility and that it is agent causation that makes agents responsible for their actions. According to Stefaan E. Cuypers, recently people such as Randolph Clarke, Timothy O’Connor and William L. Rowe have again

tried to make a case for a credible and viable agent-causal theory (Cuypers, 1998, p. 278). Cuypers discusses the idea of agent-causality in the traditional Aristotelian-Thomistic line as a possible alternative to the received view's naturalistic conception of human agency. His reason to look at this possible alternative derives immediately from the received view's failure to do justice to our intuitions concerning human agency and responsibility. Like libertarianism, this position is not mainstream (an example is Chisholm, 1964), though these theories too appear to be back on the philosophical agenda (Watson, 2003, chs. 13, 14).

Authors do not need to draw conclusions in order to establish incompatibilism (R). They can stick to explaining the mutual exclusiveness of the truth of determinism and the existence of responsible beings and understand their arguments as a contribution to a debate in which compatibilism is the dominant position. This is generally referred to as "traditional incompatibilism."

(B) We can argue for compatibilism (R) in the following four ways. Head-on:

(1) By arguing for compatibilism (F) and for the idea that freedom is a necessary and sufficient requirement for responsibility, while at the same time arguing that our interpretation of freedom is one that shows freedom to be compatible with determinism. This position is generally referred to as "traditional compatibilism." An example of this is the conditional analysis, discussed in section 4.A.

(2) By arguing against PAP, the principle according to which we are only responsible for *a* if we could have done something other than *a*. In this case, someone who is a compatibilist (R) with regard to responsibility and determinism can still be an incompatibilist (F). Recently, this position has been developed by Fischer (Fischer, 1994). The most important representatives of this position are Frankfurt (Frankfurt, 1969 and 1971) and Dworkin (Dworkin, 1988).

People can also argue for compatibilism (R) in a more indirect way. As mentioned above, I call this approach "practical compatibilism (R)." Practical compatibilism is the view that most people are (deeply) responsible for some of their actions regardless of the truth or falsity of determinism. Responsibility, in this view, is a primitive concept without which our daily practices cannot be adequately described; analysis concentrates on the conditions that we use to establish whether someone is a responsible agent. We can argue for practical compatibilism (R) in the following two ways:

(3) By arguing that people exist who sometimes choose and act freely and that this characteristic makes them responsible for (some of) their actions. In this case, a practical compatibilism (F) might be inferred: If determin-

ism is true then compatibilism (F) is also true, for responsible human beings (who are free to choose and act) exist. This position, as I show in the next chapter, is partly defended by Wolf.

(4) We can also argue that (most) people are responsible agents most of the time because they possess some general ability relevant to responsibility, although they are not able to do otherwise. In this case, a practical compatibilism (R) might be inferred: If determinism is true then compatibilism (R) must also be true, for responsible agents (in possession of this ability relevant to responsibility) exist. This position is consistent with incompatibilism (F) and, as I show in the next chapter, is partly defended by Wallace (Wallace, 1994).

7. Conclusion

In this chapter, I introduced the three main concepts involved in the (in)compatibilist discussions about responsibility: determinism, responsibility, and freedom. I discussed (in)compatibilism (F) with regard to freedom and determinism—the conditional analysis—and showed how it fails to capture the freedom that is generally believed to be a necessary or sufficient requirement for responsibility. As examples of compatibilist (R) views with regard to responsibility and determinism, I discussed the hierarchical authorization account and constraint compatibilism. I argued that these views fail to constitute compatibilism (R) in so far as deep responsibility is concerned. Deep responsibility, I argued, is linked to the existence of ultimacy and autonomy, the conditions for which appear susceptible to the skeptical challenge. I suggested that practical compatibilism might offer a new way out of this dilemma by treating responsibility as a primitive concept (I explain how this should be understood in the last chapter). Practical compatibilism and its opponent ultimacy pessimism will be the subject of the next chapter.