

## Two

# PRACTICAL COMPATIBILISM AND ULTIMACY PESSIMISM

Spinoza sagt (epist. 62) dass der durch einen Stoss in die Luft fliegende Stein, wenn er Bewusstseyn hätte, meinen würde, aus seinem eigenen Willen zu fliegen. Ich setze nur noch hinzu, dass der Stein Recht hätte.

(Spinoza says, if a stone that is flying through the air because it was kicked would be conscious, it would believe it was flying out of its own will. I only add to this that the stone would be right.)

Arthur Schopenhauer (Schopenhauer, 1859, Zweites Buch sec. 24)

### 1. Introduction

The previous chapter concluded with what I called “the skeptical challenge”: the challenge to explicate a condition that accounts for the autonomy-related distinction between deeply responsible human beings (as we normally assume ourselves to be) and non-deeply responsible ones (such as very young children and mentally or volitionally incapacitated individuals).

In this chapter, I elaborate on and discuss two positions with regard to this condition, the positions of:

- (1) the ultimacy pessimist, who believes that such a condition necessarily involves an incompatibilist notion of freedom, and
- (2) the practical compatibilist, who believes the opposite.

Both agree that the existence of robust alternative possibilities—strong freedom of the will—is incompatible with determinism, and both reject the hypothetical and conditional analysis of “could have done otherwise.”

Ultimacy pessimism is a construction of several pessimist arguments that can be found in the literature on responsibility—arguments that focus on the (im)possibility of being the ultimate source of our actions (for example, Nagel, 1979; Russell, 1992, 2000; Kane, 1996, ch. 5). Practical compatibilism is constructed as a result of and is informed by the views of Susan Wolf and R. Jay Wallace. I believe that a new and interesting line of argument is made possible by their work.

## 2. The Practical Compatibilist Position

In the debate on the justification of our daily practices of responsibility, Wolf and Wallace position themselves in between the pragmatists and the metaphysicians (Wolf, 1990, p. 17; Wallace, 1994, p. 85).

The metaphysicians assume a prior realm of facts, independent of our daily practices of responsibility, on the basis of which we should validate these practices or show them to be unjustified. If the metaphysicians reflect on the conditions that make an individual an appropriate subject of our moral sentiments, they will discuss whether these conditions presuppose a kind of freedom that is incompatible with determinism, and, if so, whether we can make sense of the idea that indeterminism is true, or of the idea that responsibility does not exist.

The pragmatists, on the other hand, reject the idea of such an independent realm of facts. They believe that the justification of our daily practices of responsibility is constituted by the actual moral sentiments that we have toward one another and the social and practical purposes that these serve. They believe that our moral sentiments are foundational. The only questions of justification with regard to responsibility are those that arise internal to our practices, leaving the practice of responsibility as a whole intact. An unhappy consequence of the pragmatist position is that it is unable to account for the (in)compatibilist (R) discussion. According to the pragmatists, incompatibilists misrepresent and misunderstand our daily practices of responsibility. According to them, the only intelligible position is the compatibilist (R) one.

Wolf and Wallace agree with the pragmatist about the importance and priority of our (reactive) attitudes and emotions, but disagree with the pragmatist that compatibilism is the only intelligible position. Wolf and Wallace argue that the issue about which (in)compatibilists (R) disagree, is a normative issue about the conditions that should regulate our daily practices. According to this normative rendering of the argument, the incompatibilist (R) believes that one of these actual conditions presupposes the existence of strong freedom of will, whereas the compatibilist (R) believes that it does not. Contrary to the pragmatists, both Wolf and Wallace believe the incompatibilist position intelligible (although they also believe it false). Let me explain this.

### A. Pragmatism

The pragmatism that Wolf and Wallace partly accept and partly reject is informed by Peter F. Strawson's well-known essay on freedom and resentment (Strawson, 1962). Strawson argues that the thesis of determinism (a thesis he claims not to understand, Strawson, 1962, p. 59), if intelligible at all, is irrelevant to our daily practices of moral responsibility. He has the following three strongly interrelated arguments for this:

- (1) We cannot abandon our daily practice of responsibility.
- (2) Even if we could abandon our practice of responsibility, we have no reason to do so.
- (3) The only reasons to withhold our reactive sentiments are the excuses and exemptions that function in our daily practice. None of the excuses—which function locally (such as physical constraints and coercion)—or exemptions—which function more globally (such as insanity or childhood)—can be generalized into a thesis that resembles the thesis of determinism (Wallace, 1994, p. 119).

Let us take a closer look at these arguments.

The first argument is a naturalist one. According to Strawson, our practices of moral responsibility are intertwined with (embedded in, and constituted by) the larger practices of ordinary interpersonal relationships in which we relate to one another as participants, as equals capable of reciprocity. Our commitment to this framework of interpersonal relationships is so natural and inevitable that we have no choice in the matter: We cannot abandon it (Strawson, 1962, pp. 68–70).

As is often pointed out, this contention is problematic in face of the existence of so-called shame cultures. In these cultures, some argue, the reactive sentiments are absent, they cannot be said to be natural and inevitable. This difficulty can be solved with the help of Wallace’s reactive account of the relation between the moral sentiments and responsibility in which, as we saw, not so much the moral sentiments with which we react to a transgression of the moral demands are central, but instead the availability of a reaction to the transgression of the moral demands (see ch. 1, sec. 3.A). Not the reactive sentiments that happen to be available in our culture are natural and inevitable, but the availability of *some* reaction to the transgression of a moral demand. What is natural to and inevitable in every human society is not how we react but that we react.

The second of Strawson’s arguments is a “rationalist” one, as Paul Russell calls it, or a “pragmatist” one, as Wallace calls it. According to it, giving up our reactive attitudes and emotions would be irrational (compare Hertzberg, 1975). Even if we had reason to repudiate our daily practices of moral responsibility, we would have overriding reason not to, for doing so could only entail that we should exclusively take the objective attitude toward one another, an attitude in which we view each other as subjects for manipulation and treatment. When we are dealing with young children or psychologically confused individuals, this attitude might perhaps be rational, but to take this attitude in general would immeasurably impoverish our lives (Strawson, 1962, pp. 74–76).

As Russell observes, Strawson’s rationalist and naturalist arguments are in conflict with each other (Russell, 1992). If we are (psychologically) incapable of abandoning the reactive sentiments, to argue that abandoning them would impoverish our lives would be pointless. This is acknowledged by

Strawson as he points out that questions concerning the rational justification of the inter-personal attitude “could seem real only to those who utterly failed to grasp the preceding answer, the fact of our natural commitment to ordinary inter-personal attitudes” (Strawson, 1962, p. 70). Despite this conflict, ultimacy pessimists and practical compatibilist both accept this and the previous naturalist argument. They disagree about the third argument, that Wallace has called the “internal” one (Wallace, 1994, p. 96).

Strawson argues that of all the excuses or exemptions, none can be generalized into a situation resembling determinism. With regard to the excuses, Strawson contends that our reactive sentiments are related to the quality of our will instead of its freedom. According to him, we react to and care about the goodwill and ill will or indifference of others toward us or toward a third party. If people hurt us but “could not do otherwise” we will cease to feel the reactive sentiment of blame or resentment, because we understand their proclaimed inability to do otherwise as an indication that they meant no harm and did not deliberately hurt us. With regard to the exemptions, Strawson argues that the considerations to suspend our reactive sentiments toward an agent either at the time of the action (if, for instance, the agent acted under severe conditions of stress) or all the time (if, for instance, the agent is psychologically incapacitated or morally undeveloped) are the same as those that invite us to take the objective attitude toward the agent. We exempt only those that qualify for treatment and manipulation.

Since neither of these reasons to withhold our reactive sentiments—the lack of ill will or the lessened accountability—can be generalized into a situation resembling determinism, the truth of determinism cannot be relevant to our practice (Strawson, 1962, p. 65). The idea behind this argument is the following.

We do not hold people responsible if some excuses or exemptions obtain. Determinism, if intelligible at all, is a general thesis. If determinism is true this would only entail that no single individual is morally responsible, if the thesis were to embody a generalization of the acknowledged excuses or exemptions. When we generalize the first type of excuse—the excuse that someone did not mean ill—we get something like a reign of universal goodwill, a situation in which nobody means ill. A situation like that, if imaginable at all, can hardly be said to be threatening to our practice of moral responsibility. The objection to the second generalization is more interesting for our purposes. According to Strawson, a generalization of the second group of excuses would make abnormality the universal condition, which is incoherent (Strawson, 1962, p. 68). This is contestable. To assert that “we normal human beings are in fact all abnormal” and that “all normal circumstances are exceptional” appears incoherent only if we do not specify what “normal” refers to.

Imagine, for instance—this is an example of Paul Russell’s—that a severe genetic mutation affects our brains in such a way that we all become morally undeveloped and mentally incapacitated (Russell, 1992, p. 299). What is nowadays called “abnormal”—being morally undeveloped and mentally inca-

pacitated—will in the genetically mutated brain situation become the normal situation. In this sense the apparently incoherent assertion that “all normal human beings are abnormal” is intelligible. Likewise, circumstances can become so exceptional, for instance due to war, that no people exist anymore who are morally responsible for their actions, that is, judged by the conditions we use in times of peace.

The appearance of paradox and incoherence disappears once we realize that “normal” refers to the “universal condition” in the same way that “here” refers to “this place.” Claiming that abnormality could become the universal condition, therefore, must be compared to claiming that “we are not always here,” using “here” to refer to the place we are at, instead of using it as an indexical (which always refers to the place we are at). If the conditions of responsibility that regulate our daily practice are not intelligible if determinism is true, then abnormality *is* the universal condition if determinism is true. In this sense, we still stand in need of an explication or specification of the conditions of responsibility. Without such explication or specification, we cannot conclude anything about the relevance of the truth of determinism for our daily practice (compare Russell, 1992, p. 301; see also Sie, 1998b). Practical compatibilists, I argue next, provide us with such an explication.

### B. The Normative Shift

Practical compatibilism elaborates on the necessary and sufficient conditions of responsibility by focusing on our ability to recognize, know, and fulfill the normative expectations that we have of one another—the expectations concerning what appropriate, good, and bad behavior is. Only when we know how we expect one another to act and behave—only when we know the relevant norms and values—do we know how to display ill will or goodwill toward one another (how to hurt or please others, how to be inconsiderate or considerate, and so on). The ability to act wrongly presupposes not so much an ability to do otherwise, but an evaluative ability, the ability to know right from wrong.

Wolf argues that our moral sentiments are related to the fact that we expect one another to do the right thing for the right reasons. Hence, according to her, our accountability is not constituted by our ability to do otherwise, but by our ability to do the right thing for the right reasons. The question whether we are deeply responsible subjects, according to Wolf, is not so much a metaphysical as a meta-ethical, and perhaps also an ethical one (Wolf, 1990, p. 71). We can only answer it once we know exactly what doing the right thing for the right reasons is. Wallace, in a similar vein, elaborates on the notion of a “culpable choice,” that in fact encompasses all the ways in which we are able not to do the right thing for the right reason or, in Strawsonian vocabulary, all the ways in which we are able to exercise or display ill will. Wallace argues that the unifying characteristic of the category of actions that we find blameworthy is the quality of the will they express. What we blame, resent, or are indignant

about is the culpable choice that some actions express. Wallace derives his account of excuses and the corresponding notion of a culpable choice from John L. Austin and Strawson (Wallace, 1994, pp. 118–127).

As we saw in the previous chapter (sec. 3.A), Wallace argues that our practices of responsibility are regulated by the moral demands (the accepted moral expectations) that we hold people to. These moral demands must in turn be understood as the obligation not to act in ways that express certain choices. Wallace illustrates this with the moral obligation of non-maleficence. According to him, this is not an obligation not to make some bodily movements that harm other people. Instead, this is the obligation not to act in ways that express the choice to harm other people, in the ordinary pursuit of your ends (Wallace, 1994, p. 128). This is convincing. Lots of bodily movements occur, intentional bodily movements, which harm other people without meriting any of the moral sentiments. We do not blame a dentist, for instance, for hurting us, although this happens in the ordinary pursuit of the dentist's ends. A dentist makes bodily movements that harm people, but these do not typically express the choice to harm them.

According to Wallace, the degree of moral fault is determined by the quality of the choices on which we act, regardless of whether we achieve what we aim to achieve. This is convincing. If someone decides to use a car despite its badly functioning brakes—thereby running the risk of causing an accident—the agent is blamable for driving the unsafe car through the city, regardless of causing an accident. The choice on which the agent acts is a bad one, and that is enough for blame.

What if the agent is not free to make a different choice from the one made? Wallace rejects the idea that the only culpable choices are genuine choices and that the only genuine choices are free choices (choices that are incompatible with the truth of determinism). According to him, those who argue that determinism would leave no room for culpable choices because it leaves no room for free choices prematurely decide the issue. For, according to Wallace, the very question that is at stake is what the conditions of responsibility are and whether they are compatible with determinism (Wallace, 1994, p. 149). It cannot be assumed that the ability (or opportunity) to make a genuine choice is a condition of responsibility and that genuine choices are excluded by determinism to start with. This would make incompatibilism (R) true from the outset.

As I explain in the next chapter my intuitions go in the opposite direction: I find it hard to understand the notion of a culpable choice that is not free in some fairly robust sense. However, I agree with Wallace that we should first examine the actual conditions of responsibility and see whether they presuppose free will. I also agree with Wallace that the (in)compatibilist debate is about the conditions that would render it appropriate to hold a person morally responsible (Wallace, 1994, p. 85). This normative interpretation is, according to Wallace, the only viable one left if we agree with Strawson that the moral sentiments are natural and inevitable, to the extent that we cannot take the

metaphysical view of our daily practices of responsibility very seriously (Wallace, 1994, ch. 4).

The widespread view in post-Strawsonian discussions about responsibility is that the metaphysicians feign amnesia to the facts as we know them. We treat one another as appropriate subjects of the moral sentiments all the time, and criticism, modification, and (questions of) justification are only intelligible from within that framework. Although Wallace agrees with this, he does not believe that Strawson's naturalist and pragmatist arguments suffice to silence the incompatibilists or pessimists. According to Wallace, someone can be an incompatibilist even if that person accepts that the moral sentiments are natural and inevitable, and that, if we could give up on them, this would only be impoverishing. He argues that our standards of justification may commit us to the view that freedom is a necessary condition of moral responsibility (Wallace, 1994, pp. 103–104; see also Russell, 1992). If they do, and if this freedom is incompatible with determinism, and if determinism is true, then our practices, although inevitable, are incoherent. What this means is that we are not able to live up to our standards of fairness.

Like Wallace, Wolf also understands her project as the examination of the residual possibility that our practices of responsibility could prove to be incoherent or unjustified internally. She wants, as she puts it, to bring out the truth that is embodied in the perspective that the discussion of responsibility is concerned with the fairness of the rules that govern our practices of responsibility, whereas the fact that we are responsible is fixed (Wolf, 1990, p. 17). I discuss how we should understand this exactly in the last chapter of this book. For now, let us understand it as the claim that the concept of responsibility is a primitive one, a concept we cannot do without, although we are not sure to what it refers. Let us also assume that even if we admit that it is primitive in this sense, the conditions that determine to whom it refers can still prove to be unfair. Let us, finally, also assume that the incompatibilist doubts and worries somehow concern this unfairness. What exactly, then, are these doubts and worries about?

### **3. Ultimacy Pessimism**

Although ultimacy pessimists share the practical compatibilist's view that our moral sentiments are somehow natural and inevitable, they also believe that: (1) in our daily practices we do not hold people responsible for actions that are not up to them but are due to things outside their control, and (2) if determinism is true, all individual agency is equally the inevitable effect of things outside our control, and none of our actions are truly up to us. If none of our actions are ultimately in our control or truly up to us if determinism is true, the ultimacy pessimist argues, then how can we hold any individual person *p* deeply responsible for an action *a*, if determinism is true? If determinism is true, bad events and actions might be tragic and wrong, wrongdoers might be

bad, evil, and wicked, but none of them will be deeply responsible for their actions, because ultimately their actions are not up to them. The fact that they add evil to this world is just their bad moral luck (compare Nagel, 1979; Russell, 2000).

Practical compatibilists agree with ultimacy pessimists that our daily practices are regulated by a condition of control and “up to us”-ness, but they argue that these conditions are unproblematic. We possess volitional, reflective, and evaluative abilities that make us responsible. Hereafter, I call these abilities the responsibility-relevant abilities, for short, the “RR abilities.” Practical compatibilists believe that the possession of, and/or the ability to use, the RR abilities can explain why some of us deserve to be blamed or praised for some of our actions whereas others do not—without the notion of strong freedom of will. Let us take a closer look at how this is supposed to work exactly. In what ways are the RR abilities involved in our practices of responsibility? Roughly, four distinctions are efficacious.

The first distinction corresponds with that between the mentally or volitionally incapacitated or immature people (very young children and the insane) and the mentally and volitionally sound and mature ones (adults and the sane). I call the condition that accounts for this distinction the “autonomy condition.” The second distinction is that between those who exercise the RR abilities and those who are temporarily or occasionally deprived of, or lack, the opportunity (or ability) to exercise the RR abilities fully. This can be due to, for instance, hypnosis or manipulation. I call the condition that accounts for this distinction the “autonomous action condition.” The third distinction is that between those who suffer from an inability to translate the output of their RR abilities into action, as is the case with kleptomaniacs, and those who do not suffer from such an inability. I call the condition that accounts for this distinction the “intentionality condition.” The fourth distinction is that between those who are able to act according to the outcome of their RR abilities and those who could not have done otherwise than they in fact did because of some external impediments to their freedom. I call the condition that accounts for this distinction the “intentional action condition.”

Ultimacy pessimists and practical compatibilists agree that the intentionality-related conditions—the conditions that are primarily concerned with what I called “superficial responsibility”—can be met with Harry G. Frankfurt’s only-because principle, the principle Frankfurt proposes to replace PAP with (see ch. 1, sec. 4.B). Ultimacy pessimism and practical compatibilism also agree that the autonomy-related conditions cannot be met by Frankfurt’s only-because principle; they disagree on what should be the alternative. Let me start with a short discussion of the intentionality-related conditions. Since practical compatibilists and ultimacy pessimists both object to the hierarchical view, let me discuss them in relation to the Frankfurt-counterexamples.

### A. The Intentionality-Related Requirements

I believe the hierarchical view plausible and sufficient with regard to the intentionality-related conditions because the notion of acting “in spite of yourself” is clear and unproblematic. We all know what it means to act in spite of yourself. To think about freedom of will as the opposite of this kind of unfreedom, therefore, connects to a familiar experience.

#### i. The Intentional Action Requirement

Consider the following Frankfurt-counterexample. Imagine Mary—by stipulation a sound, mature, and morally responsible subject—waiting for a friend in a car parked next to a canal. Although Mary does not know it, her friend has accidentally left her in a locked car (because of a childproof lock). If, at that moment, Mary were to see a child stumble and fall into the canal, she would not be able to get out of the car and save the child from drowning. Imagine two possible worlds: one in which Mary is a busy person, and one in which she is a considerate person. Busy Mary notices a child playing dangerously near the canal, but concentrates on the notes on her lap because she is eager to finish the work that she took with her. Considerate Mary, on the other hand, keeps an eye on the child, concerned that something bad might happen. In both possible worlds, the child stumbles and falls into the water and Considerate Mary and Busy Mary fail to prevent that from happening. Busy Mary fails to notice the child fall in; Considerate Mary finds herself unable to leave the car because its doors are locked.

Let me stipulate that the concept “actions” used in these examples concerns the common-sense and broadly construed class that also includes eventful omissions or failures to make the bodily movements a situation clearly requires.

Considerate Mary will be excused. She was unable to open the doors and was unable to prevent the child from falling into the water. She is not deeply responsible for what happened. This is what the only-because principle—the principle that Frankfurt replaces PAP with—states: People are not morally responsible for what they have done if they did it only because they could not have done otherwise (Frankfurt, 1969, p. 10). As a consequence, Considerate Mary should not be blamed. She acted in spite of herself: She would have done otherwise if she could have done otherwise. This enables us to formulate an intentional action condition without robust freedom of will. Mirroring the only-because principle—which determines who should be excused—we should hold people responsible if they did not act in spite of themselves (provided that no other excuse obtained). Since people sometimes do not act in spite of themselves, we know that some people act intentionally and are in this sense responsible for their actions.

## ii. The Intentionality Requirement

Determining Busy Mary's deep responsibility for events is slightly more difficult. The closed doors are not an excuse for her behavior, because they played no role in her failure to prevent the child from falling into the water.

Traditionally, a case such as that of Busy Mary is used to establish that the lack of alternative possibilities does not count as an excuse. Due to the unavoidability of the action "not saving the child," we will re-describe the action of Busy Mary as one in which she only failed to "try to save the child," but it is unlikely that it would or should alter our moral assessment of her. The moral assessment of an agent pertains to the bodily movements that the agent makes under conditions that are necessary to make it an action. Information about the external circumstances that pertain to the action's success or failure cannot be of significance for our moral assessment, because it would not tell us anything additional about the agent (Frankfurt, 1969, pp. 8–9).

The point of the example here is that the lack of *that* excuse does not establish that Busy Mary is fully responsible for failing to prevent the child from falling in. It might be that some other exemptions or mitigating circumstances obtained. Suppose, for instance, that Busy Mary is a workaholic. Suppose that she has been working hard for years and years as a consequence of which her body produces chemical substances that make her unable to relax or concentrate on things other than her work. Busy Mary the Workaholic decides to keep an eye on the child but fails to do what she decided, because she is irresistibly drawn toward her notes. Busy Mary the Workaholic would have tried to prevent the child from falling into the canal had she stuck to her decision to watch the child (this also by stipulation). Just like Considerate Mary, Busy Mary the Workaholic would have failed to prevent the child from falling into the canal because the car doors were locked. Nonetheless, the only-because principle is still applicable. Busy Mary the Workaholic acted as much in spite of herself as Considerate Mary did, although the closed doors did not obstruct her because something else prevented her from being obstructed by them.

The only-because principle, then, helps us explicate both the intentionality condition and the intentional action condition. You are excused if some so-called external impediments interfered with your ability to act in accordance with what you wish (Considerate Mary) and if some so-called internal impediment renders you unable to act as you wish to act (Busy Mary the Workaholic). (We should take care, here, with the terms "internal" and "external," for both Marys are excused for an action that is in some way external to them, external to their true or real self. Both actions derive from a source that the Marys do not acknowledge as their own. I come back to this in sec. 3.B.iii below.) If, on the other hand, you did something because you wanted, decided, or intended to, you are responsible for it—no matter what impediments to your freedom obtain. Unfortunately, the notions of "impediments" and "unfreedom" are much more complicated in relation to the autonomy-related requirements, for

we can still ask, and often do, whether Mary the Workaholic is responsible for becoming a workaholic. This brings us to the autonomy-related distinctions.

### B. The Autonomy-Related Requirements

The notion of acting “in spite of yourself,” clear and unproblematic with regard to the intentionality-related distinctions, is difficult and obscure where the autonomy-related distinctions are concerned. Can I, for instance, be said to smoke despite myself—or, perhaps, despite my true self—if I want to quit smoking? Is my mere identification with my smoking habit enough to make smoking a habit for which I am deeply responsible? What if I do not know that nicotine is addictive? Do I act despite myself if I am manipulated into doing something I would not have done without the manipulation? Am I responsible for allowing myself to be manipulated? And so on. At this point ultimacy pessimists and practical compatibilists part company because they give different answers to these questions.

#### i. The Autonomous Action Requirement

An example of an agent who acts according to the outcome of the exercise of her RR abilities would be a busy Mary who is not addicted to her work. Let us call her: “Hardworking Mary.” Hardworking Mary decides to finish the work she took home as soon as possible because that is what she is paid for, whereas keeping an eye on an unknown child is not. The difference between Busy Mary the Workaholic and Hardworking Mary is that the second’s relation to her action concentrating on her work is superficially correct, whereas Busy Mary the Workaholic’s is not. Busy Mary the Workaholic has a problem because she is unable to do what she wants or decides to do, whereas Hardworking Mary has no such obvious problem: She formed the intention that moved her all the way into action on the basis of her choice.

It is only natural to suppose that at one time or another Busy Mary the Workaholic must become aware of her failure to act as other people do and as a result seek help. If she does not, we can, and probably will, doubt that her claimed inability is a real inability. However, that is beside the point. With regard to the question whether Hardworking Mary is deeply responsible for her failure to prevent the child from falling into the water, the mere fact that she herself formed the intention to concentrate on her work is inconclusive. Her intention to concentrate on her work might have come about in a way that we normally experience as unnatural or inappropriate.

Suppose, for instance, that Mary’s decision to finish her work was provoked by extensive manipulative measures taken by her boss. Perhaps her boss, who knows that she is an insecure creature, indoctrinated her completely by suggesting for weeks and weeks that she would prove herself unfit for her job

if she did not finish the particular assignment she was working on in time; or perhaps (business interests being enormous) the boss went even further and hired someone to hypnotize her, thus provoking the firm belief that she should concentrate on her work and nothing else. According to the ultimacy pessimist, this would make us less inclined to blame Mary than we would if her action came about under normal circumstances.

The worrying point about hypnosis and manipulation—and also about, for instance, addiction and behavioral conditioning—is that they employ the very action-producing mechanisms that normally ensure that the actions we perform are ours. The boss who manipulates Mary in such a way that she becomes obsessed with her work uses knowledge of Mary and her skills to influence her in such a way that Mary is led to believe, wish, or decide things as though she wishes, wants, or decides these things on her own. If we agree with ultimacy pessimism that Manipulated Mary is not as responsible for her actions as she would have been had she acted on her own, we are on a slippery slope. When are our actions completely our own? If we grant ultimacy pessimists their worry, we must immediately answer the question why desires, volitions, and wants provoked by post-hypnotic suggestion and manipulation are any less our own than those provoked by advertisements, the influence of friends and loved ones, our basic needs, and so on—circumstances that we normally do not understand to be and do not treat as autonomy-undermining. Therefore, most people will be reluctant to grant the ultimacy pessimists their worry. I have more to say about this reluctance later on. Let me first discuss a possible solution that immediately springs to mind, a solution in terms of ignorance.

Can we not account for our intuition that we are less deeply responsible for an action *a* if it came about under circumstances of hypnosis and manipulation by arguing that in those circumstances we act, partly, out of ignorance? Mary is ignorant of the fact that she has strong feelings of uncertainty and that someone else has manipulated these. Perhaps if Mary had had all the relevant information, she would have wished, decided, or intended to do something else and, consequently, she would have done something else. We hold Hypnotized Mary less responsible for *a* than we would have done had she not been hypnotized because she was ignorant of the role that post-hypnotic suggestion played in her performance of *a*.

This does not work. Imagine, for instance, that we are ignorant of the fact that our decision to finish our work is triggered by some material wishes we have that can only be fulfilled if we continue to have an income. Would this make us less deeply responsible for our decision to finish our work? To be sure, the fact that Mary does not know why she acts as she does might in itself be a reason to blame her, if we believe that people should take responsibility for their actions—but that is another matter. The point to be made here is that our behavior can be caused by motives that we are ignorant about but which do not lessen our responsibility. Ignorance appears to lessen our responsibility only if the circumstances we are ignorant about are already understood to be autonomy-undermining. The very notion of “autonomy-undermining circum-

stances” appears to depend upon some prior account of autonomy. Hence, a condition for autonomous action cannot be stated solely in terms of the absence of ignorance.

The most successful and influential way to solve the problem of the autonomy-related distinctions is that of the hierarchical authorization account originated by Frankfurt and closely related to his arguments against PAP as discussed above and in the previous chapter (ch.1, sec. 4.B).

## ii. Hierarchical Authorization Accounts

Hierarchical authorization accounts of responsibility meet the autonomous action condition by arguing that agent  $p$ ’s authorization of an action  $a$  makes  $a$  an autonomous action of  $p$ —no matter what the circumstances in the past or present are. If Mary decides, wishes, or intends to do  $a$  and decisively identifies with the decision to do  $a$ —or “wholeheartedly wants” to do  $a$ —she is fully responsible for  $a$  (Frankfurt, 1987). In practice, Frankfurt admits, we probably never know whether or when someone was decisive or wholehearted. These concepts are not entirely clear. His point is that if anything matters to our responsibility for our actions, it is the wholeheartedness with which we undertake them (Frankfurt, 1998, p. 34).

Hierarchical accounts of autonomy are attractive to compatibilists because they are explicated solely in terms of the relation individual agents have with their bodily movements. If an agent  $p$  does  $a$ ,  $p$  is responsible for  $a$  if the relation  $p$  has with the bodily movements that constitute  $a$ —at that very moment—is of the right authoritative kind. The external and historic determinants of the  $a$  are completely irrelevant to  $p$ ’s responsibility for it (compare Fischer and Ravizza, 1994). No matter what circumstances obtain or what causal sequence preceded  $a$ , if we actively authorize it, we are as fully responsible for it as anyone can be. As a consequence, determinism is completely irrelevant to our practices of responsibility. We can make the distinction between actions for which we are deeply responsible and those for which we are not, without the notion of alternative possibilities and strong or robust freedom of the will.

The ultimacy pessimist will object that it is not clear what exactly it means to authorize our actions and for what reasons the authorizing act and ability make us responsible for our actions. Let me illustrate the ultimacy pessimist’s point with the notion of addiction and Frankfurt’s notion of decisive identification.

Suppose that Mary is living some eighty years ago and suffers from a very light form of asthma (this example derives from Sie, 1998a). Her doctor advises her to smoke a cigarette, and she, unaware of the addictive effects of nicotine, keeps on smoking them for years. In the first instance, she just keeps on smoking because she likes it and because she believes it helps relieve her asthma. After her husband and children start complaining about the smoke and

the appalling smell, she reflects carefully about the pros and cons of smoking. Although she is still unaware of the addictive effects of nicotine, she now knows that smoking does not relieve her asthma, and in fact makes it even worse. Nevertheless, she likes smoking so much that she cannot imagine herself without her cigarettes. She decides that, all things considered, she is a smoker.

Now suppose that in a parallel universe exactly the same facts obtain except that the cigarettes smoked are of a non-addictive, herbal kind. Addicted Mary and Herbal Mary both decisively identify with the continuation of their smoking habit. From a time-slice perspective, both Marys have an equal desire to continue smoking, and both are equally unaware of the source of their strong desire. If we want to argue that decisive identification is relevant to our responsibility for some of our actions in the face of this example, we should either:

- (1) assert that both Marys are equally responsible for continuing to smoke, or
- (2) allow for the distinction between a real decisive identification (of Herbal Mary) and the mere experience of a decisive identification (of Addicted Mary).

If we allow for the second possibility, we admit that we need yet another condition with which to distinguish between cases in which our decisive identification with *a* makes us truly responsible for *a* and cases in which this is not the case—which Frankfurt, for one, will not do (see Frankfurt, 1998). The ultimacy pessimist's argument with regard to this second option will then be that these conditions, in turn, must appeal to a notion of strong freedom of will and alternative possibilities.

The first assertion is highly counterintuitive and probably in contradiction with the very notion of addiction. It is difficult to see how we can make sense of the notion of “addiction” without the notion of “autonomously operating desires,” desires that have some efficacy regardless of the agent's mental activity towards them. Let me add that if we are not convinced by an example that understands addiction as an autonomy-undermining circumstance (in comparison with the consumption of non-addictive substances and regardless of the strength of the next desire for the consumption of the substance), we could perhaps imagine the existence of addictions that make the addicted individual desire something and invalidate the individual's ability to cope with this desire. This is, as we will see below, an important part of Wallace's argument in favor of the condition with which he proposes to account for the autonomy-related distinctions (this chapter, sec. 4.B).

For the moment, I will not pursue this point any further. Even if we are not convinced by these arguments against the hierarchical authorization account of the autonomy-related distinctions, practical compatibilism does not disagree with ultimacy pessimism on this score, so I allow myself some argumentative room here. The overall aim of this book, after all, is to reconstruct

the practical compatibilist's arguments against ultimacy pessimism and to evaluate the success of these arguments. When it comes down to it, I see no ultimate objection with which to refute the arguments that motivate the hierarchical accounts. I suspect that much of the attraction of hierarchical accounts derives from the attention they draw to our ability to take responsibility for our actions, which most of us regard as an extremely important trait. Not only is their account of this trait incorrect, we can do better justice to it. My expectation is that when all the arguments of this book are in their proper place, the hierarchical view will lose much of its attraction as a solution to the (in)compatibility issue (ch. 4, sec. 5). For now we should leave it at this.

The hierarchical views on responsibility do not stop the ultimacy pessimist or the practical compatibilist from worrying about cases of addiction, hypnosis, and manipulation. Both the ultimacy pessimist and the practical compatibilist insist on the need for, and the difficulty in, making a clear distinction between good (normal) and bad (autonomy-undermining) sources of our motivational make-up (desires, decisions, wants, intentions, and so on). One influential suggestion made in connection with this is to remedy these shortcomings by an appeal to a notion of a "real self."

### iii. The Real-Self View

The real-self view—discussed under this heading by Wolf—argues that agents are deeply responsible for their actions if they were able to act on the basis of their volitions and able to base their volitions on their valuational system (Wolf, 1990, ch. 2). The view was developed in reaction to some problems with the internalist character of Frankfurt's account of the correct relation of agents to their bodily movements by Gary Watson (Watson, 1975).

The ultimacy pessimist objects—correctly, I believe—that all the problems that we have in accounting for the distinction between a good source and a bad source of our motivational make-up can be repeated with regard to the distinction between a good and a bad source of our valuational system. Even if the real-self view could define exactly what an external source of motivation is—which I do not believe it can—it would have the same problems as hierarchical views have with regard to the autonomy-undermining circumstances. Real-self views are unable to account for the set of autonomy-undermining circumstances that might be involved in the formation of someone's so-called real self. If people have suffered a severely deprived childhood, for example, are they not less responsible for their behavior than we normally assume ourselves to be?

Consider, for instance, the Harris boy that Gary Watson tells us about (Watson, 1987b). The Harris boy figures in a story written by Miles Corwin about a brutal, cold-blooded criminal who murdered two school kids and finished what was left of their lunch after committing the crime. The story narrates how Harris the young sensitive child turned into Harris the brutal and

cold criminal after an extremely harsh, deprived, and violent youth. Although Watson uses the example to illustrate that Harris is no less responsible for his behavior because of his terrible youth, the ultimacy pessimist will rightly point to the obvious and disconcerting amount of moral luck in the Harris case. Most of us, the ultimacy pessimist will argue, are terribly lucky not to have been brought up like the Harris boy to become a moral monster.

Likewise disconcerting for the real-self view are the so-called *Brave New World* cases, examples of situations of what Robert Kane has called “covert non-constraining control”—situations in which everyone is free to want and choose whatever they want and choose, but are thoroughly conditioned to want and choose only some things (Kane, 1996, p. 65). The endless possibility of autonomy-undermining measures does not stop with our motivational make-up.

Of course, these cases are controversial. Not everyone believes that an abusive childhood or awkward political circumstances lessen people’s responsibility for their actions (see, for example, Watson, 1987b; Frankfurt, 1998). Then again, the disagreement about our responsibility for these cases might be due to the fact that not many people doubt that we are most of the time free to do one thing instead of another. If the belief in our robust freedom underpins the idea that we are fully responsible for our actions even if we suffered an abusive childhood or find ourselves in awkward political circumstances, it discloses a commitment to indeterminism, not—as compatibilists would have it—to compatibilism. We need not decide on these issues at this point. Practical compatibilists have no quarrel with ultimacy pessimists about the existence of autonomy-undermining circumstances, they only disagree on the question how to account for them. This, then, finally brings us to the condition that practical compatibilists and ultimacy pessimists disagree about: the principle of alternative intentions.

### C. The Principle of Alternative Intention

The ultimacy pessimist believes that the only way to meet the autonomy-related requirements is to postulate the necessity of alternative possibilities at the level of our intentions. Let us call this the “principle of alternative intentions” (hereafter: “PAI”). PAI states that a person is deeply responsible for action *a* if that person had the opportunity, or ability, to intend to do other than *a* (compare McKenna, 1997, p. 79).

An important background condition for this principle is that that for which we are responsible can be transferred backwards in the chain of events leading to the action *a*. If we manipulate ourselves into a position in which we are no longer able to intend otherwise, then we are responsible for *a* even if we had no opportunity or ability to intend otherwise, but we are so because we could have prevented ourselves from being in the situation which necessarily led to *a*. Remember our example from chapter one of the person impersonating a lifeguard (sec. 4.B.i). If the impersonating person is not responsible for the

impersonating, that person is also not responsible for the action that is the inevitable result of the impersonation. If somewhere in the causal chain leading to the action *a* the impersonating person could have intended otherwise, that person is responsible for *a*.

According to the ultimacy pessimist, PAI is the underlying principle that accounts for our acceptance of post-hypnotic suggestion, manipulation, addiction, and Brave New World cases as autonomy-undermining. What is equally missing from all these cases, the ultimacy pessimist argues, is a psychological freedom, that is, the opportunity, or ability, to intend something other than *a*. And although it might not be completely clear what exactly this opportunity or ability to intend otherwise is, some notion like it is, according to the ultimacy pessimist, efficacious in our everyday acceptance of autonomy-undermining circumstances.

The objection that, ultimately, the acceptance of PAI undermines our deep responsibility for any of our actions does not impress the ultimacy pessimists. This is exactly what they believe and what worries them. This brings us back to the reluctance we noted above (in sec. 3.B.i), the reluctance to accept some conditions for the autonomy-related distinctions because they do not suit our intuitions that we are responsible for most of what we do. As we saw in section two of this chapter, ultimacy pessimism is formulated after a commitment to the inevitability and indispensability of our daily practices of responsibility. That we are naturally inclined and rational in treating one another as responsible subjects is not what is disputed. To reject and be outraged by the ultimacy pessimist's suggestion because we are committed to the general framework of moral sentiments is not appropriate. All the ultimacy pessimist argues is that we treat one another as deeply responsible human beings although—on reflection—this appears to require a kind of freedom that is incompatible with determinism.

If determinism is true, the ultimacy pessimist argues, we are all parts of a causal chain that started long before we were even born; so, the pessimist asks, what justifies us blaming the people unlucky enough to have become one of the bad parts instead of one of the good? This is the question practical compatibilists try to answer. Let us take a look at what is needed to do so successfully.

If we want to establish the irrelevance of PAI for responsibility, we need a counterexample to it in which person *p* is (a) in possession of the RR abilities; (b) deprived of the opportunity (or unable) to exercise the RR abilities to intend otherwise than *a*; and (c) still deeply responsible for *a*. If someone is deeply responsible for an action *a* without having the opportunity or the ability to intend differently, PAI is false. Premise (a) is necessary because PAI (contrary to PAP) is meant to explicate a condition of autonomous action. As indicated in the previous chapter, the assumption that it was an autonomous human being who intended to do *a* undermines the validity of the counterexample (ch. 1, sec. 4.B.ii). We are now able to spell out exactly why.

Remember the evil scientist Black who is able to take sufficient measures to pre-empt all alternative possibilities of Jones. He does so by installing an

ingenious mechanism in Jones' brain that ensures that Jones is unable to intend to do anything other than what Black wants her to intend to do. With regard to PAI, this either deprives Jones of her RR abilities—in which case (a) does not obtain—or it does not deprive her of the ability to intend, choose, or decide differently at all, in which case (b) does not obtain (ch. 1, sec. 4.B). If, on the contrary, we already assume that Jones' relation to her action came about in an appropriate way—if we assume that Jones is a responsible human being who correctly intended to do *a*—we presuppose what we set out to establish. We presuppose that actions exist for which we are deeply responsible even if no alternative possibilities occur, whereas—in order to refute the ultimacy pessimist—we should establish that actions take place for which we are deeply responsible even if no alternative possibilities exist. I elaborate on this in the following chapter. Let us first discuss practical compatibilism's arguments against ultimacy pessimism.

As we have seen, both Wolf and Wallace recognize the ultimacy pessimist's worry, although they do not believe it appropriate. According to Wallace, PAI is invalid; he argues that we have better principles with which to account for our practices, principles that do not employ a problematic, incompatibilist notion of freedom. According to Wolf, PAI is compatible with physical determinism; PAI is only incompatible with psychological determinism, and psychological determinism is most probably untrue (compare Wallace, 2000).

#### 4. Practical Compatibilism and the Autonomy Condition

In order to refute PAI or show it to be compatible with determinism, practical compatibilism starts with the autonomy condition; it tries to account for the distinction between responsible beings and non-responsible beings by establishing that what makes us responsible is the possession of some general abilities. To establish this, practical compatibilism needs to define—to begin with—what it means to act in a responsible way. Without that, it cannot determine which abilities are relevant to responsibility, let alone that most of us possess them.

As we saw in the second section of this chapter, Wolf's formal definition of responsible behavior is "to do the right thing for the right reasons" and Wallace's is "not to act in ways that express a culpable choice." The RR ability that corresponds with this is the ability Wallace calls "the power of reflective self-control," the ability to grasp moral reasons and to translate them into action. This, I believe, provides for a very strong anti-skeptical answer. It appears undeniable that some people exist who do the right thing for the right reasons and who are able not to make culpable choices some of the time. It appears warranted to conclude that some of us possess the RR abilities to a sufficient degree, although we might still disagree about the sufficient and necessary conditions of responsibility that make it fair to hold an agent *p*

responsible for action *a*. Nevertheless, this would never lead to the conclusion that none of us is ever responsible for anything we do. Does this address the ultimacy pessimist's worries and problems? Let us start with Wolf.

#### A. Wolf

I concentrate mainly on Wolf's arguments in *Freedom within Reason* (1990). Wolf accepts the ultimacy pessimist's principle PAI in so far as it concerns our responsibility for the class of bad and wrong actions. According to her, though, a class of actions exists—good actions—for which the question whether the agent could have intended otherwise does not arise. If someone did something nice or good—say, saved a child from drowning—the question whether they could have intended otherwise is irrelevant: It does not affect the praiseworthiness of the act. If we say of someone, for instance, that “she could not hurt a fly” or that “she could not resist buying a present for her friend,” we do not thereby claim that she is any less responsible for the actions that flow from these traits (namely, never hurting anyone or buying a present) than she would have been could she have intended otherwise.

Although we can have some second thoughts about the distinction between blame and praise, Wolf's observation here is meant to establish something about the relation between the source of our motivation, its possible inevitability, and our moral sentiments. In this sense her arguments establish a counterexample to PAI. The by nature nice and good person *p* is (a) in possession of the RR abilities; (b) unable to exercise the RR abilities when confronted with the possibility to do the right thing or unable to intend anything other than to do the right thing; and (c) still deeply responsible for doing the right thing. As we have seen, when someone is deeply responsible for an action *a* without having the opportunity or the ability to intend otherwise, PAI is false.

Wolf, to explain her point, compares two swimmers: Both dive into a canal to save a child, both have the same reasons to do so, and both are rational. The only difference between them is that one of them, *p1*, is determined to do so—the outcome of *p1*'s saving the child was inevitable—whereas the other one, *p2*, could have done otherwise. According to Wolf, *p1* is just as praiseworthy as *p2* (Wolf, 1990, pp. 61–62). The only reason to assume a difference to exist between *p1* and *p2* is the mistaken idea that *p2* did something more difficult than *p1*—for instance, resist an inner struggle not to save the child—or the mistaken idea that *p1* acted mechanically, whereas *p2* acted freely.

First of all, Wolf argues, *p1* and *p2* do not differ in the degree of inner struggle they had to overcome. The reasons they had to save the child and those they had not to save the child are, by stipulation, the same. If they had to overcome an inner struggle, they had to overcome an equal inner struggle. The only difference is that the outcome of the inner struggle was determined in *p1*'s case, but not in *p2*'s case. Secondly, Wolf argues, *p1* and *p2* do not differ in so

far as the one acts mechanically—for instance, out of blind habit—and the other does not. Acting mechanically, she points out, is the opposite of rational acting. Since it is stipulated that both acted in accordance with reason, neither acted mechanically. Both did the right thing for the right reasons. Both are equally deserving of praise (Wolf, 1990, pp. 59–62).

These arguments give away Wolf's practical compatibilist assumptions, in the sense that they are only intelligible if we assume that acting in accordance with Reason is possible regardless of determinism's truth or falsity. Wolf embraces the existence of responsible acting individuals prior to, and independent of considerations concerning how exactly determinism coheres with our Reason. Wolf stipulates that acting rationally is not the same as acting mechanically, and also that two people can act exactly the same while one is determined to do so, the other not. This relies on:

- (1) a particular interpretation of determinism,
- (2) a certain understanding of our ability to reason, and on
- (3) a particular understanding of the (in)compatibilist (R) issue.

It does not allow, for instance, room for someone defending incompatibilism (R) with the aim to argue against determinism's truth. I come back to these points in the following chapter. Let me first proceed with Wolf's arguments.

Wolf's elaboration on good actions and the irrelevance of alternative possibilities is meant to establish that autonomy-undermining circumstances do not exist in so far as someone does the right thing for the right reasons. If we act in accordance with Reason, the circumstances in which we do so or came to do so are not important. Only if we act on no reasonable basis at all can we be said to act in autonomy-undermining circumstances. Wolf uses "Reason" and "reasons" to refer to anything that can lead to true or good values, including, for instance, emotions, excluding only the non-rational or unreasonable (Wolf, 1990, p. 56). If we act on no reasonable basis at all we cannot be said to do the right thing for the right reasons.

One possible objection to this line of reasoning, Wolf notes, is to deny the existence of good reasons or good things and argue that those people who act according to the dictates of reason are not deeply responsible at all, but are controlled by an illusion. As a consequence of this, Wolf continues, we should conclude that only those who act for no reason at all are responsible for their behavior. If we argue thus, however, we are radically changing the subject: we give up on the idea that praiseworthy actions exist at all—which means, Wolf argues convincingly, that we give up on the intelligibility of the idea that we can be deeply responsible for our behavior. In the true value skeptic's view nothing can be praiseworthy or blameworthy, not even acting on no reason at all. True value skepticism entails that no good or bad things exist (Wolf, 1990, p. 65).

With regard to the class of actions that are bad or wrong, matters are different. According to Wolf, if someone does the opposite of the right thing for

the right reasons, circumstances that undermine autonomy do exist. If someone acts wrongly, this person is only to blame for it if the circumstances did not prevent the person from doing otherwise, that is, from doing the right thing for the right reasons. (This allowance of autonomy undermining circumstances in an account of the conditions of responsibility makes it into what is called a “bipartite” account.) If people cannot help having the values they have, if they cannot help being the persons they are, and if they cannot help acting in the way they do, then, according to Wolf, they are not deeply responsible for acting as they do. In so far as wrongdoers are concerned, Wolf endorses a principle very similar to PAI.

Wolf’s arguments against ultimacy pessimism focus on the idea that PAI is incompatible with determinism. She argues that although PAI is necessary, it is highly unlikely that PAI itself is incompatible with determinism. PAI would be incompatible with, for instance, psychological determinism, but, Wolf argues, psychological determinism has virtually nothing to support it (Wolf, 1990, p. 101). (I am not convinced that Wolf’s remarks about psychological determinism suffice to refute ultimacy pessimism. I do not discuss my objections at this point because they receive ample attention in the following chapter.)

Having excluded psychological determinism, Wolf asks herself what else the incompatibility of PAI with determinism could possibly imply. Suppose someone possesses RR abilities, but performs a wrong action, then *iff* the circumstances of the action prevented, or interfered with, the exercise of the RR abilities the agent can be excused. Could these circumstances of action be some tiny, unseen events that prevent the agent from doing the right thing for the right reasons (Wolf 1990, pp. 102–103)? This, Wolf continues, would be a genuine case of the inability to do otherwise because of autonomy-undermining circumstances.

According to Wolf, not every version of determinism implies that something prevents us from doing something by blocking (the exercise of) our relevant capacities. Especially doubtful is that physical determinism—which she calls physiological determinism—does so. Since, as we saw in the first chapter, physiological determinism is the kind of determinism that occupies a central role in the discussion of the legitimacy of our daily practices of responsibility, this would appear to lessen the burden of proof on the compatibilist shoulders. Let us take a closer look at Wolf’s arguments concerning physiological determinism.

#### i. Compatibilism (F)

Wolf tells us what she calls “a Leibnizian story.” With this story she questions the assumption that the physical level of explanations is more basic than the psychological level, an assumption that most people take for granted in the discussion on responsibility. By “psychological level” Wolf refers to the level

of explanations in which psychological concepts and laws figure (Wolf, 1990, pp. 109–110). But why, continues Wolf, should we assume that “I go where my body goes” instead of “my body goes where I go”? As long as my psychological freedom to make some choices is not interfered with—and this is not necessarily implied by the truth of physiological determinism—we are psychologically free to choose otherwise and responsible for what we do (Wolf, 1990, pp. 103–116).

Wolf admits that the relation between physical and psychological explanations of human behavior is subject to very difficult discussions in the philosophy of mind and the philosophy of science. (For criticism of Wolf’s treatment of these issues and a suggestion remedying it with the help of the work of Donald Davidson, see Grau, 2000.) That the relation itself is controversial, Wolf argues, means that the reducibility or the elimination of the psychological level cannot be taken for granted beforehand (Wolf, 1990, p. 109). Her point is even stronger: She believes that in so far as her story is plausible, this hints at the possibility that one day the defense of a non-reductive and non-eliminative account of the psychological level will probably succeed. In either case, incompatibilism (R) needs an extra argument in order to show that determinism is incompatible with responsibility. Nothing in our daily practices suggests that we often lack the psychological freedom to do otherwise. As long as we are able to intend otherwise in the circumstances of the actions that we perform, we are responsible for those actions. I doubt whether this point is relevant in relation to the existence of blameworthy actions but leave this matter aside here, for it is the subject of the next chapter.

If—one day—physiological explanations prove to be more basic than psychological explanations, then our daily practices of responsibility will be threatened, according to Wolf. On that day, however, our daily practices of responsibility will be threatened and all our talk in terms of “choice,” “ability,” “human action,” and so on. Until such proof is provided, we have little reason to believe that incompatibilism (F) is true and even less reason to act as though it were true. With regard to the latter—the reasons we have to act as though it were true, the practical consequences of her view—she argues that in practice the determination of someone’s psychological freedom has never been an all-or-nothing matter. We determine whether someone is able to do the right thing for the right reasons by looking more closely at the individual and by comparing that person to other individuals “like them,” for example, to individuals who share the same history or the same sociocultural background, and so on. According to Wolf, ability in this sense is a matter of degree (Wolf, 1990, p. 86–88). I come back to this point in the following chapter, section 2.A and section 3.

Leaving the possible difficulties with Wolf’s claims aside for the moment, we can appreciate the point of Wolf’s account—the point that makes us discuss her under the heading of “practical compatibilism”—that the question whether an agent could have done otherwise, if relevant at all, is not a counterfactual and metaphysical one, but primarily a practical one. Circumstances that

undermine autonomy might be “abnormal” in the literal sense of the word: They deviate from what we are used to and result in behavior that deviates from what we are used to. Although this does not establish compatibilism in any traditional sense, we might say that it establishes a practical compatibilism.

It establishes that our daily practices of responsibility are fair enough, as long as we keep an open mind to possible kinds of psychologically determined behavior. If people act wrongly and appear to do so as the consequence of unfortunate circumstances, then we should be aware that they might not be responsible for their behavior. Since psychological determinism is most likely untrue, though, we may rest assured that, in general, people are free to do the right thing and, therefore, are responsible for most of what they do. If we want to enlarge our freedom and responsibility within this world, we should “promote an open and active mind and an attitude of alertness and sensitivity to the world” (Wolf, 1990, p. 147). For this is what enables us to appreciate the True and the Good and to direct our lives in accordance with that.

I hold this reinterpretation of the relevance of theoretical truths for our day-to-day practices to be one of the impressive aspects of Wolf’s view. She locates the ultimacy pessimist’s worries and doubts, allows them some room but shows us all the reasons for not giving in to their possible skeptical consequences. Sure, she says, we did not create ourselves and we did not create the world, and in this sense we are not responsible for how responsible we are. Nevertheless, we do not need or want a freedom from the world, we want a freedom within the world. This is exactly what our possession of Reason gives us: the ability to act in accordance with the True and the Good. As long as determinism does not necessarily interfere with that ability—and it does not appear to do so, for, Wolf says, we act in accordance with reason all the time—we have little reason (and surely no overriding one) to doubt the fairness of our daily practices of responsibility.

Attractive as Wolf’s Reason view may be in its general outline and inventive arguments against the idea that determinism necessarily threatens to undermine our day-to-day practices of responsibility, it has an obvious drawback. It addresses the ultimacy pessimist’s worries and doubts by telling a story about the possibility of a world in which the pessimist’s worries and doubts are unwarranted—granted, a world that arguably looks much like ours. But the ultimacy pessimist need not doubt the possibility of a world in which pessimists’ doubts are unwarranted. The ultimacy pessimist worries about this world (compare Fischer and Ravizza, 1992, p. 386). The ultimacy pessimist will not be convinced by Wolf’s arguments against determinism’s relevance. The pessimist might grant Wolf her conclusion that determinism does not necessarily undermine our daily practices of responsibility, and continue to argue that in our world it appears very likely that determinism does undermine our practices.

The pessimist might argue, for instance, that the more we learn about the complexities of our brains and the numerous ways in which it can be impaired, the more likely it becomes that wrong actions are the inevitable product of some physiological deviations. I come back to these ultimacy pessimist’s wor-

ries later on. Let us first proceed with Wallace's answer to the ultimacy pessimist. Wallace, contrary to Wolf, does not believe that PAI is valid.

#### B. Wallace's Replacement of PAI

In *Responsibility and the Moral Sentiments*, Wallace argues that only two valid so-called principles of fairness are accepted in our practices of responsibility: the principle of "no blameworthiness without fault" (Wallace, 1994, pp. 135–147) and the principle "that it is unreasonable to hold someone to expectations that she is incapable of fulfilling" (Wallace, 1994, p. 161). The principle of "no blameworthiness without fault" could function as the condition for the intentionality-related distinctions in a way similar to Frankfurt's only-because principle. According to Wallace, a person *p* is deeply responsible for her or his bodily movements *iff* they constitute an action *a* that expresses some culpable choice made by *p*. This condition describes all possible excuses that we accept in our daily practices. If we excuse one of the Marys from the previous section for her actions, we do so not because she lacked alternative possibilities, but because she did not make any moral fault in the first place: Mary is excused for her bodily movements *iff* these bodily movements did not express a culpable choice.

With regard to autonomy-related distinctions, Wallace argues that it is unreasonable to hold someone to a moral obligation if that person lacks the power to grasp and comply with the reasons that support the obligation. People, according to Wallace, are not deeply responsible for their actions in general if they do not possess the RR abilities or if their RR abilities are somehow impaired. Likewise, people are exempted from responsibility for some actions if circumstances or conditions temporarily impair their RR abilities. Autonomy-undermining circumstances are those circumstances that affect either our ability to grasp moral reasons or our ability to respond to them appropriately, that is, to control our behavior in the light of these reasons.

This account is very powerful. For one thing, it enables us understand and illuminate the difficult and complex discussions and disagreements about the meaning and nature of addiction in relation to responsibility (see sec. 3.B.ii above). Wallace discusses Frankfurt's notion of a "willing addict" (Frankfurt, 1971, pp. 24–25) as follows.

According to Frankfurt it is possible that a severely addicted individual *p* is responsible for the actions that stem from *p*'s addiction. If *p* wants, decisively identifies with, or decides to act as *p* does, *p* is fully responsible for it. According to Wallace, the correctness of this conclusion depends on the way in which addiction functions. If an addiction affects an agent's capacity for practical reasoning itself, the mere fact that *p* decisively identified with doing *a* does not make it reasonable to hold *p* deeply responsible for *a*. *P*'s decisive identification, after all, came into existence because of a lack of the reflective capacity for self-control. The source of all *p*'s desires is *p*'s addiction and the

source of what these desires lead *p* to intend and consequently to do, is also *p*'s addiction.

If, on the other hand, addiction functions only episodically as a source of desires, the addict is capable of reflective self-control some of the time. *P* is able to choose between taking measures against the addiction—although they should perhaps be “Odysseus-like”—or identifying with the addictive situation (Wallace, 1994, pp. 172–174). (“Odysseus-like” refers to the possibility of taking measures against yourself.) In such cases it would appear reasonable to hold the addict deeply responsible for an action that flows from an addiction. If the addicted person *p* is in possession of the reflective powers of self-control, and if *p*'s action *a* expresses a culpable choice, then *p* is deeply responsible for *a*.

This latter example appears to constitute a counterexample to PAI. The addicted person whose addiction functions only episodically as a source of desires is: (a) in possession of the RR abilities; (b) unable to exercise the RR abilities to intend anything other than to take the drug when confronted with the substance, and (c) still deeply responsible for taking the drug. PAI, on Wallace's account, appears false.

Wallace claims that what holds for addiction also holds for the other possible circumstances that undermine autonomy, such as manipulation, post-hypnotic suggestion, aversive conditioning, and so on. If circumstances or conditions do not undermine an agent's power of reflective self-control, then a manipulated, hypnotized, conditioned, or drugged person is no different from a non-manipulated, non-hypnotized, unconditioned, or undrugged person. After all, says Wallace, we are all burdened with desires and volitions that have their source outside ourselves; we are all influenced by friends, advertisements, and so on (Wallace, 1994, p. 175).

I do not believe that this argument, as it stands, succeeds in refuting ultimacy pessimism, for that is exactly what worries ultimacy pessimists: that that kind of external influence will eventually erode all our moral judgments. Since I elaborate on the ultimacy pessimist's argument in the next chapter, I do not go into this point here.

The above assertion might be taken to indicate Wallace's practical compatibilist assumption that no matter what theoretical truths we uncover, deeply responsible individuals and normal responsible behavior do exist.

Straightforward influences, then, do not diminish our status as deeply responsible subjects, although Wallace admits to the ultimacy pessimist that there could be a slippery slope from straightforward influences down to more autonomy-undermining ones, such as subliminal advertising and systematic deception (Wallace 1994, p. 175n). These—like all other autonomy-undermining circumstances—undermine our autonomy by invalidating our reflective power of self-control. This is crucial to Wallace's argument: autonomy-undermining circumstances or conditions must disable the RR abilities themselves in order to be exempting at all. With this restriction—that only those circumstances that temporarily invalidate our RR abilities can be accepted as

exempting circumstances—Wallace explicitly rejects what he calls “a bipartite account” of the conditions of responsibility (Wallace, 1994, p. 208). According to such an account, people need the RR abilities and the opportunity to exercise this general ability in the particular circumstances of action. Here the paths of Wolf and Wallace diverge. Wolf, as we have seen, endorses such a bipartite account (sec. 4, this chapter).

According to Wallace, a bipartite account strengthens the principle of fairness that the incompatibilists often use to voice their incompatibilist worries: the “principle of avoidability.” The principle of avoidability holds that people should possess general RR abilities and should also be able to avoid the moral wrong that they are held deeply responsible for. Besides possessing the general power of reflective self-control, an agent who is held responsible for a moral wrong also needs the opportunity to exercise these abilities in the circumstances of action. This, subsequently, is argued to be incompatible with determinism.

The only version of this argument that Wallace accepts is the version that holds that cases of, for instance, addiction or the effects of stress should not be described as cases in which the general powers of reflective self-control are impaired, but as cases in which it is hard or very difficult to exercise them. It appears, he admits, unfair to hold someone deeply responsible if they can grasp and follow some moral reasons only with extraordinary difficulty, a difficulty that borders on the impossible. This impossibility, though, is not the impossibility that figures in the thesis of determinism. It is perhaps impossible for us to do anything other than what we did, if determinism is true, but this kind of impossible is temporally restricted, whereas the kind of impossibility that exempts is not.

If determinism is true and at moment  $t_1$  I did not do a push-up, then it is physically impossible for me to do a push-up at  $t_1$  even if I proceed to do twenty push-ups after  $t_1$ . The impossibility that exempts someone from responsibility, on the other hand, is not temporally restricted like this. People are exempted from responsibility for  $a$  if it is impossible for them to perform  $a$  in the sense that it lies beyond their general powers. The latter sense of “impossible” is related to the natural laws and facts about some individual’s constitution and capacities, whereas the sense of “impossible” that flows from determinism is related to the laws of nature and facts about the past (Wallace, 1994, pp. 215–217).

To conclude, according to Wallace both the intentionality-related and the autonomy-related conditions can be explicated without presupposing the necessity of alternative possibilities or the necessity of the possibility to intend otherwise. PAI, therefore, is as irrelevant to our daily practices as PAP is. As long as people act on the choices that they make and as long as these choices sometimes violate the moral obligations that we accept, people are responsible for the actions on the basis of these culpable choices. The only reason to relieve people from bearing the responsibility for the culpable choices expressed in their actions is a temporary or more permanent impairment of

their RR abilities. That determinism would somehow undermine our general powers of reflective self-control (Wallace, 1994, p. 199) is hard to believe. Why would people who possess the reflective powers of self-control and are able to avoid the harm imposed by moral sanction in daily life, become any less deeply responsible for their actions in a deterministic universe?

Like Wolf, Wallace transfers to the incompatibilist (R) the burden of proof with regard to the legitimacy of our daily practices of responsibility. Our practices of responsibility can be accounted for without alternative possibilities, and as such they appear perfectly reasonable and legitimate, and determinism appears quite irrelevant. Contrary to Wolf's view, Wallace's view addresses the ultimacy pessimists' worries and doubts by arguing against their principle of alternative intentions. In doing so, he develops a powerful account of the reason why excuses excuse—that is, because they show that no moral fault was made—and a powerful account of why exemptions exempt. Since we are responsible *iff* we possess the RR abilities, we are not responsible if these RR abilities are absent or temporarily impaired.

With these accounts of the excuses and exemptions, Wallace successfully refines Strawson's internal argument against the pessimist. He develops conditions to meet the intentionality and autonomy-related distinctions, while at the same time exploring in what ways these conditions mention anything possibly absent if determinism is true. In the case of the excuses, the important items are culpable choices; in the case of the exemptions, the central item is our power of reflective self-control. Since both culpable choices and our power of reflective self-control, according to Wallace, are intelligible without alternative possibilities, we have no reason to accept incompatibilism (R).

Wallace's view also has a slight drawback: He does not illuminate what it means exactly to possess the ability of reflective self-control. Although I accept that we do not stand in need of such an illumination as long as we discuss the existence of responsibility in terms of our ability to act in accordance with the moral demands we accept, we do stand in need of such an explication if we are confronted with a so-called exemplary wrong action. In what terms should we phrase the idea that someone possesses the capacity for reflective self-control when that person does not do the right thing for the right reasons? Would it be wrong for us to say that the person who acted wrongly, could have done otherwise? And how can people who possess the power of reflective self-control act in completely wrong ways? These questions are the subject of the following chapter.

## **5. Conclusion**

Both Wallace and Wolf establish compatibilism by giving an account of our daily practices of responsibility and the distinctions that obtain, without allowing a notion of freedom that is incompatible with determinism to enter the picture. On top of this, they argue that determinism is not (necessarily) incom-

patible with the overall assumption of these practices that most of us possess the RR abilities and act accordingly most of the time. This allows for the conclusion that, as long as we have no reasons to doubt the fairness of our practices of responsibility, we might as well assume that they are compatible with determinism: If determinism proves to be true, then compatibilism must be true—a conclusion, as might be clear, that depends on the prior commitment to us being deeply responsible for some of the things we do.

The condition that both Wolf and Wallace explicate with regard to the skeptical challenge is that people are deeply responsible if they possess some abilities that enable them to know what they should do, and to act according to this knowledge. Because we know that people exist who know what to do and act in accordance with that knowledge, doubt or pessimism regarding our practices of responsibility appears unwarranted. So, we can stop ultimacy pessimism in its tracks—or can we?

As we saw in section three above, the ultimacy pessimist believes that if determinism is true no distinctive blame- or praiseworthy activity exists that justifies our blaming or praising that agent. With regard to praiseworthy actions, we might grant that practical compatibilism establishes that a distinctive activity on the part of well-acting individuals exists for which we could praise them: they know what they ought to do and do exactly that. According to Wolf, we can praise such persons for this because to recognize that some actions, characters, and lives are better than others, to see which ones are better than others, and to control your behavior so as to make your actions, character, and life better instead of worse is praiseworthy (Wolf, 1990, p. 77). To mirror this, we could say that wrongly-acting agents must be to blame because they know what they are doing, know that they should not do it, but do exactly that and do so without any good excuse or exemption. They are to blame for their wrong actions, because they know which actions, characters, and lives are better than others, but do not control their behavior so as to make their actions, character, and lives better instead of worse. It remains to be seen, as I argue in the following chapter, whether this idea can be made intelligible without reintroducing PAI.