CHAPTER 5

OVERCOMING SCEPTICISM? BELIEF AND MORAL RESPONSIBILITY

Carlos J. Moya

As a result of the preceding chapters, we now have a very strong case for scepticism about moral responsibility. In this last chapter, we shall try to see whether this scepticism could ultimately be avoided. Even if it could, however, the strength of the case for it should already lead us to suspect that the scope of moral responsibility, in the sense of true desert, might be narrower than we acritically tend to assume, and make us more cautious about too indiscriminate and profuse applications of this concept. Even if some persons, sometimes, truly deserve blame for their actions, it may well be that the cases in which they do are actually fewer, and their blameworthiness less, than we naturally tend to think. Those who seriously follow the paths of scepticism are likely to develop a more tolerant stance towards the weaknesses of human beings and a reluctance to emit rash judgements about them. If we go deeply enough into scepticism, this will not leave us in the state of mind we were in at the beginning of our journey, even if we are finally able to overcome the sceptical predicament. I think this is a healthy consequence that can improve our lives and bring serenity and peace into them.

It is good that going through the sceptical paths can have such healthy consequences; for it may well be that there is no decisive refutation of scepticism about moral responsibility. I certainly cannot commit myself to offering such a refutation. But I can offer some rather detailed reasons for thinking that scepticism may not be the last word about moral responsibility. Let us draw the general lines of our proposal, starting with an attempt to diagnose the roots of this scepticism.

SCEPTICISM: DIAGNOSIS AND OUTLINE OF AN ANTI-SCEPTICAL PROPOSAL

The main argument for scepticism about moral responsibility, SMR, is logically valid. If we want to deny its conclusion, we must deny at least one of its premises. We have seen two lines of argument in favour of SMR's premise B, that is, the incompatibility between determinism and moral responsibility. The first line leads to premise B through the necessity of alternative possibilities for moral responsibility and the incompatibility of determinism and alternative possibilities. The second line leads to premise B through the necessity of ultimate control for moral responsibility and the incompatibility between determinism and ultimate control. The soundness of either line of argument is sufficient to establish SMR's premise B. This premise is, so to speak, logically overdetermined. However, concerning SMR's premise C, namely the incompatibility between indeterminism and moral responsibility, there is no argument that leads to its truth on the basis of the necessity of alternative possibilities, given that they are clearly compatible with indeterminism. The main line of argument that leads to SMR's premise C goes through the necessity of ultimate control for moral responsibility and the incompatibility between indeterminism and ultimate control. This line of argument, as we see, has one premise, namely the necessity of ultimate control for moral responsibility, in common with the second line of argument for SMR's premise B. Combining these two lines, we get a unified argument for scepticism, as follows:

ultimate control is necessary for moral responsibility; ultimate control is incompatible with determinism and also with indeterminism; therefore (on the assumption that either determinism or indeterminism must be true) moral responsibility is not possible. To fill in the details of this argument, remember that, in the preceding chapter, we distinguished two aspects or components of this condition, namely ultimacy of origin or source and rational control. Suppose that determinism holds. Then, whether or not rational control can be made sense of, ultimacy of source cannot be satisfied, for there are no ultimate causes or origins in a deterministic world. Suppose that indeterminism holds. Then we can have events that, being undetermined, can play the role of ultimate causes, but now the problem, as we have seen, is that it seems that these ultimate causes cannot be under the agent's rational control. In either case, it seems that the requirement of ultimate control cannot be met.

This dialectical situation suggests that the requirement of ultimate control plays a central role in supporting scepticism about moral responsibility. If ultimate control is incompatible with both determinism and indeterminism and is required for moral responsibility, the sceptical conclusion follows. This incompatibility suggests that something may be wrong with this condition. And in fact several authors have argued that the condition raises an impossible demand. If this claim is true, no wonder it cannot be satisfied, whether determinism is true or not. An important argument for the impossibility of ultimate control was developed, some years ago, by Galen Strawson. Strawson uses the expressions 'true responsibility' or 'true self-determination', but the content of these expressions is arguably equivalent to Kane's Ultimate Responsibility and to what we have called 'ultimate control'. Kane himself writes that 'Strawson's "true responsibility" [is] what I designate as "ultimate responsibility"' (Kane 1996: 34.) A consideration of Strawson's argument will be of some help in elaborating a more detailed diagnosis of the roots of scepticism. Strawson's argument runs as follows:

> (1) Interested in free action, we are particularly, even if not exclusively, interested in rational actions (i.e., actions performed for reasons), and wish to show that such actions are or can be free actions. (2) How one acts when one acts rationally (i.e., for a reason) is, necessarily, a function of, or determined by, how one is, mentally speaking ... (3) If, therefore, one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking – in certain respects, at least. (4) But to be truly responsible for how one is, mentally speaking, in certain respects, one must have chosen [consciously and explicitly] to be the way one is, mentally speaking, in certain respects … (5) But one cannot really be said to choose, in a conscious, reasoned fashion, to be the way one is, mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, '$\mathbf{P}_1$' – with preferences, values, pro-attitudes, ideals, whatever – in the light of which one chooses how to be. (6) But then to be truly responsible on account of having chosen to be the way one is, mentally speaking, in certain respects, one must be truly responsible for one's having *these* principles of choice $\mathbf{P}_1$. (7) But for this to be so one must have chosen them, in a reasoned, conscious fashion. (8) But for this, i.e. (7), to be so one must already have had some principles of choice, $\mathbf{P}_2$, in the light of which one chose $\mathbf{P}_1$. (9) And so on.

(Strawson 1986: 28–9)

Strawson's conclusion is that 'true self-determination is logically impossible because it requires the actual completion of an infinite regress of choices of principles of choice' (Strawson 1986: 29.) The ultimacy requirement that we found in Kane is clearly echoed in premises (2) and (3) of Strawson's argument: a subject's rational actions are explained by her mental constitution, or character; so, if a subject is to be truly (ultimately) responsible for the way she acts, she has to be truly (ultimately) responsible for this mental constitution, or character, by having *chosen* it. The rational control requirement appears in this argument, from premise (4) onwards, as a condition of true (ultimate) responsibility: to be truly responsible for one's mental constitution one has to have chosen that mental constitution in a *rational* way, that is, in the light of principles of choice. The argument is highly instructive. It shows why those two aspects of ultimate control, namely ultimacy of source and rational control, might not be jointly satisfied. We can make a rational choice of a certain action on the basis of our reasons, but the choice cannot at the same time be an ultimate source of the action, because it is a function of what reasons we have and find convincing and this, in turn, depends on our character, on how we are, 'mentally speaking'. But the same applies to a rational choice of our character. On the other hand, we can choose a way of acting, or a way of being, in such a way that the choice is an ultimate, underived source of such an action or character, but the choice is bound to be made on no basis at all and so in a completely arbitrary way: it cannot be a rational choice. It seems, then, that no choice can be both an ultimate and a rational source of one's actions. In the end, we are bound to act on the basis of factors that we cannot have rationally chosen and for which we cannot be truly responsible. Another way of putting Strawson's point about the impossibility of actually completing an infinite series of choices of principles of choice is to say that ultimate control involves a self-defeating demand for self-creation. As Randolph Clarke has put the point, according to Strawson 'rational free action would be possible only for an agent who was *causa sui*' (Clarke 1997: 37; cf. also Levy 2001.) Ultimate control (true responsibility, ultimate responsibility, true self-determination) is a requirement that cannot be met.

But if ultimate control is so plainly impossible to satisfy, is it not unreasonable to raise it to the status of a requirement for moral responsibility? Strawson voices a related objection to his contention on behalf of an opponent to it: 'It may be objected that the kind of freedom this argument shows to be impossible is so obviously impossible that it is not even worth considering' (Strawson 1986: 30.) And his reply is that 'the kind of freedom that it is an argument against is just the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess' (Strawson 1986: 30). We shall come back to this reply. Whether or not it is finally true, there is certainly something to it. We can develop this point in a slightly different way. The requirement of ultimate control arises out of the very nature and scope of moral responsibility attributions. From an objective, third-person point of view, a serious attribution of moral responsibility is directed to the agent herself, on the assumption that she is the true origin of the action, or consequence thereof, for which she is held responsible. These attributions have a deep and strong effect on our worth and value, as well as our self-esteem and sense of dignity. It is understandable, then, that we want to keep personal control over the grounds on which such attributions are made. Otherwise, we might find our worth increased or diminished without our participation, in quite unexpected and hazardous ways. This is why we want to have, and think we actually have, ultimate control over those things for which we are rightly held responsible. In fact, as we have already pointed out, this desire for, and belief in, control over the responsibility we bear for what we do is also at the root of the alternative possibilities condition. If we are to be morally responsible for a certain action, we understandably want to have freedom to do that action or do something else instead, including simply not doing that action. If this is not the case, it seems that we lose control over our moral responsibility, for then we might be morally responsible for actions that we could not avoid performing. This common root of the ultimate control and the alternative possibilities requirements for moral responsibility leads naturally to unifying them, as Kane does in his notion of ultimate responsibility. A useful label for this unified requirement, inspired by Fischer's terminology, which we have already used, is 'ultimate regulative control'.

These reflections speak up for the existence of a constitutive link between moral responsibility and some deep, ultimate form of control. We have argued in favour of such a link against compatibilist denials of it. Without ultimate control, it seems, we can still give some sense to the expression 'moral responsibility', but not the crucial sense of true desert, of true praise- and blameworthiness which, being a central aspect of our intuitive, pre-theoretical idea of moral responsibility, is also (and thereby) a central concern of philosophical reflection about it. Without the assumption of some form of deep, ultimate control over our actions and choices, attributions of moral responsibility, of moral praise- or blameworthiness, would come close to aesthetic judgements of people on the basis of their innate physical characteristics, such as their beauty or ugliness, or of some of their psychological capacities, such as their intelligence or their talent for music.

However, to acknowledge the necessity of a deep, ultimate form of control as central to our intuitions about moral responsibility, understood as true desert, still leaves it open what the right content of that condition is. Accepting the intuition that some form of deep control over one's actions is a requirement for moral responsibility does not imply accepting a particular theoretical construal of that requirement. It is not obvious, then, that Strawson is right when he holds that his 'true responsibility' coincides with 'the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess', or, in other words, with our intuitions about deep, ultimate control as a condition of moral responsibility. It may well be that deep, ultimate control looks impossible in virtue of a particular theoretical construal of such a condition which, on closer inspection, might be shown not to reflect our intuitions about the requirements for moral responsibility. It may well be that a philosophical construal of the idea of ultimate control is available which respects our intuitions about the necessity of that condition for moral responsibility while, at the same time, avoiding the traits of current philosophical proposals that make ultimate control appear to be an impossible demand. We shall argue that such a construal may actually be available and indicate the direction in which it might be framed. This will constitute the positive, curative part of our proposal. Before that, however, we still need to get a more precise diagnosis of the roots of scepticism about moral responsibility. We still need to investigate whether our everyday intuitions about the conditions of moral responsibility, 'the kind of freedom that most people ordinarily and unreflectively suppose themselves to possess', actually raise an impossible demand or whether this impossibility is rather the artefact of particular theoretical accounts of those intuitions. Our conjecture is that the latter is true, but not the former, and that what makes a deep, ultimate form of control appear impossible in the context of these theoretical accounts is *their almost exclusive emphasis on the will and will-related acts, particularly choices*. Our proposal, to anticipate, will be that ultimate control should be construed in terms of beliefs rather than choices, and that, with such a construal, the condition might actually be met.

This almost exclusive emphasis on the will and will-related conative acts, especially choices, as the core of ultimate control can be clearly seen in current philosophical approaches to this condition, such as Robert Kane's or Galen Strawson's. Think, for example, of the crucial role that self-forming willings play in Kane's notion of ultimate responsibility. For him, an agent can only be said to have ultimate responsibility (control) over her actions by virtue of her performing causally undetermined, plural rational and voluntary choices (self-forming willings) through which she originally builds up and brings about her own character and motives, the springs of those actions. So the very root of ultimate control over one's actions is an act of will, a self-forming willing or choice. In Strawson's case, the crucial role of choice is also undeniable. True responsibility for one's actions requires true responsibility for the mental constitution they arise from, and this, according to Strawson, implies that one has *chosen* that mental constitution as well as the principles on which such a choice is made. It seems, then, to be an unexamined assumption of Kane's and Strawson's approaches that there is a constitutive link between ultimate control and the will: it is this will-centred view of ultimate control that, in the end, justifies considering an agent as the genuine, ultimate author of her actions, thereby grounding moral responsibility attributions.

Now we shall contend that this assumption of a constitutive link between deep, ultimate control and the will, in the form of willings or choices, unlike the assumption of a constitutive link between responsibility and deep control, is not backed by our intuitions about when an agent truly deserves praise and blame for some of her deeds. I think that we are prepared to acknowledge that an agent is the ultimate source or origin, the true author and creator of a certain performance of hers, so that she fully, unrestrictedly deserves praise or blame for it, even if we do not see that performance, in any important sense, as a result of the agent's choices or acts of will. If this is true, not all control, even ultimate, is dependent on choices or acts of will. And we strongly suspect that the opposite assumption might be at the root of scepticism about the possibility of moral responsibility. Suppose, in fact, that any item over which an agent has ultimate control has to depend, in the end, on a choice of hers. And add to this the completely plausible assumption that, for it to ground the agent's responsibility for that item, the choice has to be rational: it has to be made on the basis of reasons or principles that justify and explain it. If so, according to the first assumption, she has to have chosen those reasons or principles, which in turn, on the second assumption, implies that she must already have another set of reasons or principles for the choice... and we have started the infinite regress of choices of principles of choice that Strawson rightly claims to be impossible to complete. However, if we drop the first assumption, the regress does not need to start. In fact, we shall argue that, in many cases in which we acknowledge an agent's full authorship, and so full praise- and blameworthiness, for an accomplishment of hers, it is precisely the fact that she has *not* chosen either the principles on which such an accomplishment depends or many other factors that have made it possible that makes us see the agent as truly deserving our praise (or blame) for it.

Much of the reflection on free will and moral responsibility has been based on two related, and questionable, assumptions. We have already indicated the first: it is that the will, in the form of conative acts, is the last foundation of moral responsibility. We have found this assumption in Kane and Strawson, but its tracks can be followed in many other authors. An enormous weight has been placed on the instant of decision or choice, with a corresponding oblivion of the cognitive context in which choices are made. The picture of responsible agents as constantly making decisions is quite likely distorted. I agree with Linda Zagzebski when she writes that ‘choice even in the realm of action is much less important and occurs much less often than is commonly believed’ (Zagzebski 2000b: 212). Most of the time we act on the basis of our beliefs, of our cognitive view of things, which includes our evaluative beliefs about what is valuable and worth pursuing, with no conscious decision. And in cases in which decision is involved, when the instant of decision comes, much of the matter is already settled, on the basis of such cognitive views. In contrast, it is instructive to consider the view of choice that derives from Kane’s conception of self-forming willings in the context of a response to Frankfurt cases (cf. Kane 1996: 192). According to Kane, in a Frankfurt case, a counterfactual intervener (Black, say) would not be able to control the agent’s choice provided that this is a self-forming action or willing. The reason is that, no matter how the process prior to the choice goes, Black will not be able to predict which final decision will make. Watching the agent’s deliberation process and collecting a good deal of information about it, Black may be confident that the agent will finally decide in the way he wants her to and so allow the agent to make the choice fully on herself. The agent's choice, however, may contravene Black's expectations. Alternatively, Black can wait and see what the agent actually chooses, but then it will be 'too late to control the choice' (Kane 1996: 192). This picture of choice corresponds, we think, to the assumption we are referring to. Choice, as the ultimate foundation of moral responsibility, appears to arise as something unpredictable, as an event that suddenly breaks the process of deliberation, leaving a would-be controller with no crack in which to insert his wedge. It is clear how this picture can easily fall prey to versions of the 'Mind' argument. As we have suggested, it is this view of moral responsibility as ultimately dependent on choice that makes ultimate responsibility or control appear to be an impossible demand. In our response to Frankfurt's attack on PAP, we were careful to avoid such a picture, insisting instead, against Pereboom and Zagzebski, on the importance of cognitive possibilities, such as thinking of certain moral, evaluative reasons, and on the constant availability of alternative actions even in the presence of a Frankfurtian, counterfactual controller.

The second assumption we referred to above follows naturally from the preceding one. It can be seen most clearly in Strawson's claim that, in order for true responsibility to be possible, an agent has to have chosen the principles on which she makes her choices. Unless she has chosen those principles, she cannot be truly praise- or blameworthy for any of her actions. This claim reveals a deeply individualistic view of human agents as radically self-made, self-contained entities, whose constitution does not owe anything to factors external to them, or at least it cannot do so except at the cost of their not truly deserving praise and blame. Nothing short of absolute, radical origination, which includes the choice of one's own reasons, can be enough for ultimate control, and so for moral responsibility. So, if the explanation of a person's acts traces back to factors external to her self and so to her possibilities of choice, her moral responsibility for those acts is jeopardized. On this deeply individualistic view of human agents, the fact that they appear to be socially constituted is seen as threatening to their moral responsibility. The compatibilist reaction to this is to claim that agents can be morally responsible even if their actions' origins go far beyond the agents themselves. The incompatibilist reaction is that they cannot. Libertarians, then, try to show how it is possible for an agent to be a radical, absolute source of her actions in spite of social influences on her character and motives. Non-libertarian incompatibilists hold that, given that human beings are socially constituted, the radical origination that would be required for moral responsibility is not possible at all. But the right reaction, which we shall be arguing for, is that an agent can have ultimate control of her actions and be her deep, ultimate source partly by virtue of being socially constituted. In other words, the social nature of human beings is best viewed as an enabling factor rather than an obstacle for her moral praise- and blameworthiness.

In the light of all these considerations, we can now outline the essentials of our positive proposal. Our general position can be boldly stated as follows. Ultimate regulative control is the freedom-relevant necessary and sufficient condition of moral responsibility, understood as true desert. We have defended the alternative possibilities condition (the 'regulative' aspect of control) against several criticisms of it. We have also argued for the ultimacy aspect of control against compatibilist attempts to reject it.

But we have not rejected important compatibilist insights about (rational) control. In its attempt to construe the notion of moral responsibility without resorting either to alternative possibilities or to ultimacy, compatibilist reflection on moral responsibility has brought to light other aspects of that notion which might otherwise have gone unnoticed. Following Kane, we have accepted the necessity of some of those aspects for moral responsibility. However, we have insisted that regulative control (alternative possibilities) and ultimate control are also required in order to have a sufficient condition. This contention puts our proposal on the incompatibilist side. If some form of the Consequence Argument is sound, alternative possibilities are not compatible with determinism, though they clearly seem to be compatible with indeterminism. The hardest problem, of course, comes from ultimate control. Our purpose is to show that ultimate control is indeed possible and that, even if it is not compatible with determinism, it may nevertheless be compatible with indeterminism. We shall therefore try to reject the incompatibility between ultimate control and indeterminism. This incompatibility, as the reader will recall, is an essential step in the main argument that supports SMR's premise C, namely that, if determinism is not true, moral responsibility is not possible. On this basis, we shall try to reject SMR's premise C so as to avoid the sceptical conclusion of this argument.

In order to accomplish this anti-sceptical program we shall be contending that, though the sort of control required for moral responsibility may be voluntary control, or control based on decisions or choices, it need not be. Not all control is voluntary control. The opposite assumption has led some writers to sever the link between responsibility and control, which we think is constitutive and non-negotiable. So, for example, in a recent book (Owens 2000), David Owens has claimed, on the basis that we do not have voluntary control over our beliefs, that the notion of responsibility, as well as related deontological notions, applies only to actions, and not to beliefs. We shall argue that this view is not correct and that, though we agree that belief is not voluntary, we can rightly be praised or blamed for our beliefs, for we have over them a form of control that, though not essentially based on the will, is nonetheless fit to support praise- and blameworthiness attributions.[1]

As we anticipated, instead of a will-centred account of moral responsibility we recommend a belief-centred perspective. At the foundational root of moral responsibility we shall not place conative phenomena, such as choices, decisions or (first- or second-order) volitions, but cognitive ones. Especially important will be a distinctive class of beliefs, namely evaluative beliefs about what is really valuable and worth pursuing or avoiding in life. Beliefs of this sort play a central role in explaining those of our decisions and choices that have moral import. We recommend, then, a cognitive rather than a conative conception of moral responsibility.

This cognitive turn, as it might be called, in the theory of free will and moral responsibility, is not without precedents. We already know some of them. A step in this direction is Gary Watson's emphasis on values, versus Frankfurt's second-order volitions. A more detailed contribution is Susan Wolf's Reason View, according to which the sort of freedom that is relevant to moral responsibility is the ability to act on one's values *and to form those values* in the light of an appreciation of the True and the Good. Against Wolf, however, we shall contend that ultimate control, which she finds intuitively appealing as a requirement for moral responsibility though in the end impossible, is instead possible, and also compatible with a cognitive approach, based on evaluative beliefs, to moral responsibility. So, unlike Wolf (and Watson), we shall develop our account in a libertarian direction rather than in a compatibilist one. Also significant, though not dealt with in the previous chapters, are Paul Benson's and Henry Richardson's contributions (Benson 1994, Richardson 2001), in which they insist on a substantive approach to freedom and autonomy, with an emphasis on the *content* of the agents' attitudes, against purely formal or structural approaches.[2] In fact, we can also find this emphasis on content in Wolf's conception, in which she reacts against Watson's formal view of values and Frankfurt's structural perspective.

If we refuse to place conative acts, such as decisions or choices, at the roots of moral responsibility, we also deny the view that moral responsibility ultimately rests upon conative states, such as desires. We reject a Humean view of motivation, which we have seen at work in Kane's libertarian theory, with damaging effects on his project. In this context, we find some of Derek Parfit's anti-Humean remarks in one of his papers quite suggestive and congenial. He speaks, e.g., of 'truths about what is worth achieving, or preventing' (Parfit 1997: 129), which are quite close to our evaluative beliefs. He also writes that 'the most important reasons are not merely, or mainly, reasons for acting. They are also reasons for having the desires on which we act. These are reasons to want some thing, for its own sake, which are provided by facts about this thing. Such reasons we can call *value-based*' (Parfit 1997: 127–8). In fact, if Humean internalism about motivation is true, then deep, ultimate control over our actions does not seem possible, for, even if I could form correct evaluative beliefs, they would remain motivationally inert unless they connected with a previously existent desire, a mere fact about me for which I am not praise- or blameworthy. It is important, then, that evaluative beliefs are able to motivate and give rise to desires for which they provide reasons.

Another field of research that may prove useful to our proposal is epistemology. Over the last two decades, some epistemologists have proposed looking at ethics in order to throw light on normative epistemological notions, such as justification. Roughly, a belief is justified just in case it has been formed as it *ought* to have been, where the 'ought' is the normative 'ought' that we also apply to actions. If this points to the right direction, then there is room for talk about an ethics of belief, and about responsibility, praise- and blameworthiness, in the epistemic and not only in the practical realm. The discussion about this research field is lively and contributions to it have increased enormously. In fact, however, given the strength of the case for scepticism about free will and moral responsibility about actions and the obscurity that surrounds these notions, one might be tempted to think that epistemology is unlikely to get more light about its central notions by looking in this direction. But maybe the right reaction is to encourage interdisciplinary research. Our proposal, in fact, can be seen as somehow the reversal of that general epistemological project, in that we intend to see whether scepticism about moral responsibility for our actions could be overcome, or at least undermined, by focusing on beliefs and the responsibility we bear for them.

Finally, the cognitive approach to moral responsibility that we are suggesting should also avoid the individualistic view of human agents that goes with the will-centred, conative approach. At this point, we also find interesting parallels with some epistemological issues. Cartesianism is driven by the impulse toward an ultimate rational examination and control over anything that is relevant to the process of inquiry, as a requirement of epistemic responsibility (cf. Hookway 1994: 214–15). This impulse is quite analogous to that which drives will-centred conceptions of ultimate control as a condition of moral responsibility. According to these conceptions, we cannot have such ultimate control, and so moral responsibility, for our actions unless the springs of those actions are the result of our own choice. It can be seen that both approaches adopt a profoundly individualistic stance toward human beings as cognitive or practical subjects. No epistemically or practically valuable result can be credited to an individual if the explanation of such a result traces back to factors beyond the reach of the individual's powers of rational reflection and choice. An absolute origin in the individual as a self-sufficient, self-contained entity is required of any accomplishment for which she truly deserves praise or blame. The individual may be subject to cognitive and practical influences from her social environment, but she has to make those influences part of herself by means of her own rational reflection and choice if she is to deserve credit for any result that such influences may help to explain. The problem, of course, is that rational reflection and choice require reasons and principles that, in the end, cannot be subject to rational reflection and choice except at the cost of an infinite regress. In both cases, scepticism (about knowledge or about moral responsibility) is the expected result. The question, once again, is whether we can retain the intuitions behind the ultimate control requirement without the costs of will-centred, individualistic construals of that requirement. And a positive look at the social nature of human agents, for what concerns the possibility of their moral responsibility, would seem to be part of such an achievement. We should take seriously the view which Wilhelm Dilthey expressed a long time ago when he said that a human being is a crossing-point of the large systems of social interaction. Of course, Dilthey is not alone in his insistence on the social constitution of the individual. He writes in what Tyler Burge (Burge 1979) has called the 'Hegelian' (as opposed to the Cartesian) tradition, with its emphasis on the role of social objectivities and institutions in the shaping of the individual mind. In this context, it seems right to say that the dominant lines of reflection on free will and moral responsibility have followed the Cartesian, individualistic tradition. Just as, in the philosophy of mind, Burge has favoured a Hegelian perspective on the nature of intentional content against the dominant Cartesian approach, we tentatively recommend such a perspective in the reflection on moral responsibility as well.

Let us now substantiate and develop further the proposal we have just outlined in this section.

BELIEF, CONTROL AND RESPONSIBILITY

Let us start by noting that talk about responsibility for one's beliefs not only makes perfect sense, but is quite common in everyday life. Think of such remarks as the following, which implies blame: 'How could you believe what he told you? Don't you know how often he lies?' In cases like this, similarly to what happens in the sphere of actions, excuses are likely to be expected: 'Well, this time he really looked sincere.' Other cases concern formation of belief in the light of clearly insufficient evidence. We sometimes blame people, including ourselves, for being too rash and careless in coming to have certain beliefs. On the positive side, we also praise people for not being too credulous. There are also interesting cases of praising someone for not believing something which has almost overwhelming evidence in its favour. Some thrillers provide good examples of this: think of the police inspector who keeps investigating, against her superiors' criterion, when everything seems to point to someone as the culprit of the crime. Depending on many circumstances, we tend to see some of these cases as exemplifying either the virtue of tenacity or the vice of stubbornness.

These are just a few examples of our application of responsibility-related concepts to the cognitive, and not only to the practical, field. How are we to understand this application? Some years ago, Bernard Williams (1973) famously argued that beliefs are not under our direct voluntary control, so that there is not much room for deciding to believe. It is incompatible to have a belief and to know that one has that belief only because one has decided to have it. The main reason for this contention is that belief constitutively aims at truth. As Williams writes: 'If I could acquire a belief at will, I could acquire it whether it was true or not; moreover I would know that I could acquire it whether it was true or not. If in full consciousness I could will to acquire a "belief" irrespective of its truth, it is unclear that before the event I can seriously think of it as a belief, i.e. as something purporting to represent reality' (Williams 1973: 148).

Williams's rejection of doxastic voluntarism has gathered wide acceptance. I find it convincing too. There are some disagreements, however. In a recent paper (Shah 2002), Nishi Shah contends that, if Williams's argument against doxastic voluntarism were sound, then no activity that has a constitutive aim could be performed at will. Lying, for instance, whose constitutive aim is to deceive, could not be voluntary, which is plainly false. But this criticism is mistaken. Williams's argument does not rely on the assumption that believing has a constitutive aim, but on the *particular* constitutive aim it has, namely *truth*. Beliefs aim to conform themselves to the way things actually are, whereas, e.g., decisions and desires aim to change the world so that it conforms to them. In other words, what puts beliefs beyond the reach of the will is that they have a mind-to-world direction of fit, whereas decisions and desires have a world-to-mind direction of fit. Shah is wrong, then, in holding that 'if Williams's argument goes through for the case of belief, then it will go through for any aim-constituted activity' (Shah 2002: 438).

In the context of the present research, it is good that doxastic voluntarism is false. For suppose that it is true, so that believing is, after all, an action or activity. Then, it is hard to see how appealing to beliefs could be of help in the task of overcoming scepticism about moral responsibility for our actions, for beliefs would then be affected by all the sceptical considerations that we have gone through concerning moral responsibility for actions. Inquiry is a cluster of activities: gathering and assessing evidence, drawing conclusions, etc. And it certainly seems to be true that we praise or blame people for the way they conduct their inquiries, presumably because the way they conduct them increases, or decreases, as the case may be, the probability of having true and justified beliefs. This, however, is a particular case of ascribing responsibility for actions. If responsibility for beliefs derives *entirely* from responsibility for our cognitive activity, we may expect little help from focusing on beliefs for what concerns the possibility of overcoming scepticism about moral responsibility for our actions. If there are sceptical doubts about control and responsibility for actions, the doubts will also extend to those actions related to inquiry and belief formation. So the thesis that the control we may have over our beliefs derives entirely from voluntary control over our cognitive activity is not of much help to our anti-sceptical project. But we might have a different form of control over our beliefs. This control would not be voluntary, but also not merely indirect. It would not derive entirely from our voluntary control over our epistemic activity. My suggestion is that we actually have a control of that sort, which grounds some attributions of responsibility for our beliefs.

As a first step towards a characterization of this sort of control, let us think of someone who is carrying out a rather complicated addition without an electronic calculator. She performs the task carefully and gets the right result. She is praiseworthy on both accounts, and has had control over both the process and its result. It is not a matter of luck that she has got it right. But think what having control amounts to in this case. It does not have to do with choices or acts of will in any important sense. The control she has consists rather in her yielding to the internal structure of the thing itself, the figures and the addition rules. It is, so to speak, a passive form of control, which she exercises precisely in being guided by what is there, in the addition problem. She does not choose the rules. In fact, she would *lose* control of the process if she chose the rules (or the figures), and she would rightly be blamed if she did that. She has neither chosen nor created either the rules or the figures. They come ‘from outside’ her self or her will. But this does not exclude her having deep control over her belief about the result of the addition.

This example involves an algorithmic procedure to arrive at the correct belief about the result of an addition. But we can consider more complex cases of belief and belief formation, though still with no specific moral profile. Think of great achievements in the fields of science or philosophy. They can harmlessly be taken to be systems of beliefs. These cases show, with special clarity, that authorship concerning beliefs may be as important for a person's worth and self-esteem as authorship concerning actions or decisions. Think of the virulence of many discussions, in the past and the present, about the paternity of a certain idea or theory. To mention just a couple of cases, remember the famous dispute between Newton and Leibniz concerning the invention of infinitesimal calculus, or, more recently, the discussion about the true discoverer of the virus causing AIDS. These disputes go deep into questions of personal worth and value. In connection with this, think also of the strongly negative moral assessment that plagiarism raises in most of us. Questions of real source or origin, and of corresponding praise- and blameworthiness, are no less significant and pressing in the cognitive field than in the practical one.

Let us now reflect on a particular example, Newton's *Philosophiae Naturalis Principia Mathematica*, in order to discern some aspects of our notion of authorship or source in relation to cognitive achievements. We agree, I hope, that, provided that he wrote it, Newton truly deserves our unrestricted praise and gratitude for producing such a great scientific work, which has deeply influenced our view of the natural world. We are grateful to him not only for the strenuous effort he expended in producing his work, not only for the careful cognitive activity he carried out in order to produce it, but also for the result itself, for the important ideas and beliefs that such an activity and effort brought about. He has genuine merit and truly deserves praise for his great intellectual achievement. But now consider how many elements that were arguably indispensable for his work to be possible were not of his own making, how much his work is actually indebted to cognitive factors that he did not give origin to. It is hard to see, for example, how the *Principia Mathematica* could have been produced without the contributions of such thinkers as Copernicus, Galilei or Kepler, whose work is in turn indebted to prior scientists and philosophers. And something similar can be said about many methodological and mathematical instruments and empirical data that Newton did not create or discover himself, but found already there. Nevertheless, this does not incline us to question Newton's full authorship, merit and responsibility for his work. But it seems to show that our judgements about authorship, praiseworthiness and responsibility for cognitive achievements do not correspond to the pattern of Kane's or Strawson's conceptions of ultimate control, at least for what concerns the ultimacy of source aspect. If we dig deep into the origins of Newton's work, we shall probably find a rather messy situation, with some aspects ultimately traceable to Newton himself and many others whose roots go far beyond him. But that does not prevent us from seeing his work as a whole, as an articulated system of propositions and laws, as truly and ultimately attributable to him. This suggests that something similar might be the case in the practical realm.

But I think that Newton's example can also make Kane's and Strawson's conceptions of ultimate control wanting for what regards the rational control aspect. Though surely the will, in the form of choices, is involved in the process of creation of the *Principia Mathematica*, Newton's rational control over such a process does not mainly consist in voluntary acts or choices, but, to a large extent, in his passively yielding to the internal requirements and structure of the subject itself, in his sensitivity to the actual relations of deductive and inductive justification, to the internal connections between concepts, to the perceived necessity of certain steps in the reasoning process. In fact, for him to be legitimately praised for his work, it is centrally important that he produced it with due humility and respect for the data and the principles of valid reasoning, as well as for their relations. If we discovered that he made these aspects depend on his will, our judgement about his merit and praiseworthiness would be drastically affected. This suggests that the link between the control required for true desert and the will, at least in the cognitive realm, is much looser than has traditionally been assumed in the reflection about moral responsibility.

These examples allow some tentative but fairly interesting remarks about certain varieties of control related to true desert.

First, if a particular theoretical construal of the sort of control required for true desert, such as Kane's or Strawson's notion of ultimate control, leads to rejecting the subjects' praiseworthiness for their beliefs in the preceding examples, we should conclude that, at least in the cognitive field, that theoretical construal is not correct, not that our subjects are not praiseworthy, for our intuitive judgement about their praiseworthiness is much firmer than our confidence in such theoretical constructions. They do indeed have all the control over their beliefs that is required for them to truly deserve praise for having them. And if we call all the control required for true desert 'ultimate control', they certainly have ultimate control over their beliefs. In the addition example, it would be ludicrous to deny that the agent really deserves praise on the basis that she has not chosen or created the rules of addition, or the figures she has worked with. And the same applies to many aspects of Newton's great work.

Second, although there is some distance from the preceding remarks about true desert and control over *beliefs* to conclusions about true desert and moral responsibility for actions and the possibility of overcoming scepticism concerning them, the examples also offer some clues as to how this distance might be bridged. For simplicity, let us focus on the addition example. The subject's belief about the result of the addition is a state. But her performing the calculation is certainly a sequence of actions. It is cognitive activity aimed at getting the true result of the addition. As it happens, our judgement is that the subject also deserves praise for the way she performs this activity. For the sake of argument, let us call each step in the process of calculation a 'choice'. In this case, that the subject has rational control over these choices means that she makes them according to the rules of addition. These are, in Strawson's terms, her 'principles of choice'. Now, if Strawson were right about the control required for true desert, then, in order for our subject to truly deserve praise for her calculation, she should have chosen these principles of choice. But this seems simply wrong. These principles are *the* principles to apply to this activity. So, if our subject is truly praiseworthy for performing her task carefully, according to these principles, this would seem to provide a prima facie counterexample to Strawson's construal of ultimate control as true responsibility or true self-determination, as well as to Kane's construal if, as he himself claims, his 'ultimate responsibility' is what Strawson calls 'true responsibility'. The counterexample is still only prima facie in that the case at hand does not have a specific moral import. So it might still be that *moral* praise- and blameworthiness for actions does require ultimate control in Strawson’s or Kane’s sense. But the example certainly casts some doubts on the correctness of these conceptions of ultimate control as necessary for moral responsibility.

Third, the examples also offer a prima facie case for thinking that the ‘external’ origin of the factors that explain a certain accomplishment or activity of a subject does not, in itself, provide a reason for holding that the subject is not the true origin of such an accomplishment. In the addition case, both the arithmetic rules and the figures that our subject has to work with have an external origin and are certainly explanatory of both her calculation activity and her belief. As for the rules, she has acquired them through a learning process. She has not chosen or invented them. With due respect for the higher complexity of the case, similar remarks may also be made concerning Newton’s example. But if, as we have suggested, we think of an individual as at least partly socially constituted, we do not need to deny that our subjects are the true origin of their activity and their beliefs, so depriving them of the merit they deserve. Departing from a rampant individualism and accepting the social nature of human beings provides a way of reconciling the influence of external factors on a subject’s performances and her having ultimate control over them. However, since there are cases in which external factors do actually undermine the subject’s ultimate control and responsibility (think, for instance, of Brave New World cases), this leaves open the task of distinguishing between these two possible effects of external factors on the subject’s control and responsibility, and of explaining the difference.

Finally, it seems to me that the assumption that our subjects had alternative possibilities (regulative control) underlies our judgement that they are praiseworthy for their activity and the beliefs they arrived at through it. In the addition case, it seems true that the agent could have acted less carefully in adding and arrived at a false or less justified belief about the result of her problem, and that this thought partly explains our disposition to praise her. And the same goes for the example of Newton. But we shall return to this important point in a later section.

The preceding considerations about the nature of the control required for praiseworthiness, which suggest its rather loose connection with the will, can also be seen to apply, *mutatis mutandis*, to the field of literary creation. Even if, in this field, the role of the will and choice may be much larger than in the case of scientific and philosophical creations, we still find control to depend here, to a wide extent, on the author's respect for the internal structure and tendencies of the fictional world she has created. Let us sustain this claim by arguing that one of the features that distinguish great literary narrative or drama from second-rate creations is the author's yielding to the intrinsic, autonomous dynamics of the characters, instead of constantly making decisions on their behalf about what they are to think or do at each moment. Voluntary interference in the life of the characters leads, somewhat paradoxically, to loss of control over the work, so giving rise to the impression of capriciousness and lack of depth that characterizes weak literary creations. On the other hand, and also somehow paradoxically, allowing oneself to be controlled by the intrinsic dynamics of the work is one sign of the author's control over it. Wilhelm Dilthey spoke about the experience of being led in creation, thus referring, as I interpret him, to this yielding to the internal life of the created world that underlies valuable literary works.

The variety of control we are pointing to is not a purely passive stance, however; it is, rather, a form of enlightened humility and respect for the object, which disposes the subject to recognize the proper value of what is other than herself. Far from being at odds with authorship and responsibility, this form of control is a mark of great authors and a legitimate ground for praiseworthiness in the field of literary creation. With due attention to the differences, something similar can be held to obtain, as we have seen, in the field of scientific and philosophical creation. We shall contend that it is also centrally important in the practical field, in connection with moral responsibility for actions. It is now time to come back to this problem.

EVALUATIVE BELIEFS AND MORAL RESPONSIBILITY

We have insisted that the intuition behind the idea that ultimate control is a requirement for moral responsibility, understood in the basic sense of true desert, is largely correct. We have expressed doubts, however, about the correctness of prevailing theoretical elaborations of that intuition in terms of acts of will or choices. The crucial problem with the attempt to ground moral responsibility in acts of will or choices is that, in the end, such choices are bound to be made without reasons and so in a baseless, arbitrary way, as Strawson’s sceptical argument makes clear. Our proposal is to analyse the notion of ultimate control from a cognitive perspective, in terms of beliefs. However, interested as we are in moral responsibility for our decisions and actions, evaluative beliefs about what is important, worth pursuing and caring about in one's life are especially relevant. Let us develop and justify this proposal further.

Evaluative beliefs are beliefs with an evaluative content. Not any beliefs of this sort, however, are relevant for purposes of grounding moral responsibility. Their evaluative content should express the way a person conceives of a human life that is worth living and should have potential consequences as a criterion for choice and a guide for action. In other words, it should have a moral import. So evaluative beliefs about, say, the flavour of a particular dish or about a certain car model are not of this sort, unless eating this particular dish or possessing this car are among the highest ends in a person's life (which, by the way, is not unthinkable nowadays). Given their role in the conduct of one's life, our real evaluative beliefs are likely to manifest themselves not only in our sincere general thoughts or declarations about our values but also in our particular judgements about concrete situations and in our actual choices and actions. But a straightforward logical behaviourism should be rejected, since there can be a real tension between our sincerely held evaluations and our actual behaviour, as well as between our deep, long-term values and our passing desires and preferences. Our particular choices and actions are not always a secure criterion of our evaluative beliefs.

If evaluative beliefs are to play a central role in grounding the possibility of ultimate regulative control over our choices and actions, and so of our moral responsibility for them, they have to satisfy a number of conditions: 1) Corresponding to the depth of moral responsibility ascriptions, they should be a central component of a person as a potentially morally responsible agent. 2) They should be correctly attributed to an agent as their true author and origin, in order for some of her choices and actions to be truly ascribable to her as their source. 3) The agent should have rational control over those beliefs; they should be justified and based on reasons. This requirement, however, should not give rise to a self-defeating infinite regress. 4) They should be potentially efficacious in our behaviour, even if we can sometimes act against them. 5) Corresponding to the regulative aspect of ultimate control, we should have alternatives with respect to them. 6) The justification condition (condition three) should hold even if the beliefs are not causally determined; in other words, our proposal should not fall prey to some version or other of the ‘Mind’ argument.

In this section, we shall comment on the first four conditions, while leaving the two remaining conditions to the next two sections. However, we should warn that the conditions we have just listed are not fully independent of each other: what each condition involves should be understood in connection with the rest. So a general perspective of what ultimate regulative control amounts to in the context of our cognitive proposal is only to be expected if each aspect is viewed in the context provided by the others.

Let us start with the first condition. Throughout this book, we have insisted repeatedly that ascriptions of moral responsibility have a characteristic depth. In making such ascriptions, we are blaming or praising, and so valuing, the *person* for what she has done, as we view the action for which we hold *her* responsible as an expression of her self. If this is not the case, so that we see an action as a rather hasty and accidental event, and not really as an expression of some deep trait in the person, we tend to withdraw or significantly soften our judgement. Even Hume, whom Frankfurt rightly criticizes for having dealt only with freedom of action rather than freedom of the will, acknowledges this depth of moral responsibility ascriptions when he writes that ‘actions are objects of our moral sentiment, so far only as they are indications of the internal character, passions, and affections’ (Hume 1975: 99). And it is also the perception of this depth that has led thinkers, on both the compatibilist and the incompatibilist side, to look for the roots of moral responsibility in long-term features of the agent’s self or character. Frankfurt’s second-order volitions or Watson’s valuational system are clear examples. Our proposal, centred on a subject’s evaluative beliefs, honours this insight as well. Our evaluative beliefs are a central part of how we are, mentally speaking. They inform not only some particular decisions but also the general way in which we conduct our lives and our relations to other people. They are essential to the sense we can make of our life and of our place in the world. No psychological description of a person could be minimally complete that did not include her views about what she finds important and worth pursuing or avoiding in her life. And, if we reflect on our serious ascriptions of moral praise- and blameworthiness, we shall find ourselves ultimately praising or blaming an agent for her evaluative convictions. We seriously praise or blame someone for an action in so far as we see this action as a sign of, or as flowing from, her evaluative attitudes. So focusing on such attitudes as the ground of moral responsibility ascriptions does certainly account for the characteristic depth of such ascriptions as judgements about the person, her worth and value.

Let us move on to the second condition. According to the intuition of ultimate control as a requirement for moral responsibility, if we blame or praise a person for an action of hers in so far as we see this action as flowing from her evaluative views, then, for this judgement to be justified, it has to be the case that this person is the author of, and responsible for, these evaluative views. We consider the agent as praise- or blameworthy not only for her action, but also for her evaluative convictions, for we see her action as an expression of these convictions. As we pointed out in the preceding paragraph, if we do not, so that her action appears to us as expressing a rather momentary impulse, quite unrelated to her evaluative perspective, our judgement is significantly softened, or even withdrawn, depending on several circumstances of the particular case. Our judgement of a person on account of her evaluative attitudes is not like our judgement of her on account of her intelligence or beauty. It seems that we are assuming that, unlike the latter properties, she has proper power or control over them, so that they are truly attributable to her. It is not, we think, that she has simply happened to have such attitudes, as she has happened to be intelligent or beautiful. We assume, then, that she is responsible for that part of her self that consists in such evaluations. In fact, if this assumption proves to be false, so that we cease to consider the person as the true origin of her evaluative views, as may well happen in CNC manipulation or Brave New World cases, we also stop viewing her as a proper target for moral responsibility ascriptions. But, then, under what conditions can a subject be held to be the author of her evaluative views? A look at the process through which we acquire such views will be useful in order to answer this question.

We acquire and shape our evaluative beliefs more or less in the same way as we acquire and shape our beliefs about matters of fact. Some natural, innate capacities are required in both cases. Especially important for forming correct evaluative beliefs, including moral beliefs, is the capacity to put oneself in the place of others, to try to see things from the perspective of fellow human beings. This seems essential in order to grasp the idea of a moral duty. Some subjects do not have this capacity, and this excludes them from moral responsibility ascriptions, given that they are not able to form correct evaluative moral beliefs. Moreover, as in the case of factual beliefs, we receive many of our initial evaluative beliefs from our social environment. We are initially told and taught which actions are good or bad, which ends are valuable or worthless, which attitudes toward oneself or others we ought or ought not to adopt, and so on. From the perspective we have now reached, this ‘external’ origin of our evaluative beliefs should not lead us to conclude, as is done all too often, that we are not truly praise- or blameworthy for the evaluative beliefs we end up having, and for the actions that we perform in the light of them. If that conclusion were allowed, one should also hold, to go back to our example, that Newton is not truly praiseworthy for his theoretical work, since the sources of his first (and many of his later) beliefs about the physical world were also ‘external’ to him. In fact, it is hard to see how we could have beliefs at all without external sources, including our social environment. Again, if human beings are constitutively social, it is as social beings that they can be morally responsible, morally praise- or blameworthy agents. The mere fact that we are constitutively social beings cannot be consistently used to deny our moral responsibility, for it is only as social beings that we can have moral duties and be morally responsible; but this does not imply that particular social conditions cannot undermine, destroy or promote, as the case may be, the moral responsibility of people subject to those conditions.

Now, if we really could do nothing but have the evaluative beliefs that we happen to acquire from our social environment, we would not be truly responsible for having them, and moral responsibility for our actions would not be possible. But we do not only acquire beliefs, either evaluative or merely factual; we also learn ways of forming, evaluating, accepting and rejecting them. In fact, in the absence of the latter, it is hard to see how someone could be said to have a system of beliefs. The most basic of these ways is related to the truth-aiming nature of belief itself. It consists in allowing our beliefs to be determined by how things actually are. We have referred to a higher phase of this attitude, in relation to our examples in the preceding section, as respect and humility towards the structure of the things themselves. And we have also seen how this attitude of active passivity, as it might be described in a slightly paradoxical fashion, is not incompatible with true authorship and responsibility. Another central way of assessing beliefs that we acquire is to consider their logical relations. Contradiction is especially important in this respect: two contradictory beliefs cannot both be true. Again, this basic skill may be possessed in different degrees and is certainly perfectible. Let us stop here and refer to epistemology to get a more detailed view of cognitive assessment. Now, I think we use these basic ways of assessment not only with regard to our factual beliefs but also with regard to our evaluative ones. And this use is central to the possibility of moving on from a set of merely acquired evaluative beliefs to a system of evaluative beliefs that can be said to be the agent's own and for which she can be responsible. Though with a rather different aim in mind, I agree with Henry Richardson when, dealing with autonomy, he writes:

> Although Kant's contrast between acting on a desire and acting on principle is useful to understanding autonomy, he was wrong to think of the capacity of autonomy as existing within us a priori. Rather, it is a fragile and contingent achievement. It must, first of all, be achieved. Developing the capacity to act on a conception of reasons requires learning from experience. It requires being able to articulate one's reasons and subject them to testing.
>
> (Richardson 2001: 296)

If Richardson's 'testing' is enlarged so as to include not only empirical but also logical testing, his view comes quite close to our own. Not all views about evaluative facts that we receive from our social environment need to be consistent with one another, and this sets us the task of establishing which of two contradictory views is actually true. And, even in the absence of contradiction, we may also find that a received evaluative belief is not actually true: we may discover, from our own experience, that something is not really valuable and worthwhile even if we were told that it was.

Through the application of her capacity to assess evaluative views on the basis of logical criticism and experience (though with important qualifications to be made below about the sort of experience involved), a subject may progressively advance from purely received beliefs to a set of evaluative beliefs that can be said to be her own and for which she truly deserves moral praise or blame. In a rather modest way, as compared with such great intellectual achievements as Newton's *Principia Mathematica*, but not in a completely different sense, a system of evaluative beliefs can also be attributable to an agent as its author, and thus be a legitimate source of her praise- or blameworthiness. And, in so far as this system is truly hers, she can also be praised or blamed for acting in agreement or disagreement with it.

However, as we pointed out, a fuller account of this important condition of authorship, as a central component of the general requirement of ultimate regulative control on moral responsibility, is not independent of the other conditions we are dealing with, and especially of the two last conditions we listed, which will be dealt with in the next two sections of this chapter.

Something similar should be said about the third condition, which concerns the rational control and justification of our evaluative beliefs, with particular attention to the threat of an infinite regress: though we shall start commenting on this condition now, these comments should not be viewed in isolation. The fifth section of this chapter will be especially relevant to a proper understanding of it.

Let us first see whether the sort of infinite regress that makes true responsibility impossible, according to Strawson, threatens the rational control we might have over our evaluative beliefs. These beliefs are certainly a central component of ‘the way we are, mentally speaking’, as Strawson puts it. But if Strawson’s argument shows that we cannot be truly responsible for the way we are, mentally speaking, does it not also show that we cannot be truly responsible for our evaluative beliefs? Paraphrasing Strawson, if we are to have ultimate control over our actions, we must have ultimate control over our evaluative beliefs; this means that we must have chosen them rationally in the light of certain principles of choice (which include further evaluative beliefs), which in turn we must have chosen in the light of further principles, and so on. Consequently, ultimate control is impossible.

However, from a cognitive perspective on moral responsibility, this argument can be resisted. Strawson's argument assumes that the only basis of true responsibility (ultimate control) is rational choice. But this assumption is mistaken. Concerning a certain set of beliefs, it is not true that, in order to have ultimate control over them, that is, in order to be their true origin or author and to have rational control over them, one must have *chosen* that set in the light of a further set that one must also have chosen in the light of yet another set, and so on. Remember the example of the addition in the preceding section. The subject has arrived at a belief about the result of the addition by carefully applying the appropriate arithmetical rules. She is rationally justified in holding that belief, which she herself has given rise to. She has all the control over such a belief that is required for her being praiseworthy for having it. But she has not chosen the arithmetical rules by applying which she has reached her belief. It seems, then, that choosing the principles on which we form our beliefs is not required for having ultimate control over them. Similar considerations apply to the example of Newton. He did not choose the principles of inductive and deductive reasoning that he applied in his work, nor did he choose the empirical data that he worked with. In fact, as we argued, it is centrally important for his deserved praise that he did not choose either. And the same holds for the addition example.

It is the existence of a form of rational control which is not based on choice that places our proposal beyond the reach of Strawson's sceptical argument. Rational control over our beliefs has less to do with choice or voluntary production than with forming those beliefs in the right attitude of respect for how things actually are. And the same holds for our evaluative beliefs. We exercise rational and voluntary control over our actions by choosing them in the light of such beliefs, but the control we should have over those beliefs in order to have ultimate control over our actions does not derive from a further choice of those beliefs in the light of other beliefs. Instead, as in the case of non-evaluative beliefs, it has to do with appropriate respect for the facts they purport to capture, namely facts about what is really valuable and worth pursuing in human life. Just as it is important, in order to achieve our ends, that we have true, or at least justified beliefs about matters of fact, in order to have a life worth living it is centrally important that we have true or at least justified beliefs about which ends are really valuable and worth pursuing. Otherwise we may find that achieving our ends actually does nothing to promote our happiness and wellbeing.

However, even if our proposal is not threatened by a regress of choices, it may be affected by a different regress, a regress of reasons for believing. This is, however, part of a general epistemological problem, the problem of justification. It is a problem for everybody. But perhaps focusing on evaluative beliefs may throw some light on it. As we have suggested, experience plays a central role in the formation and assessment of our beliefs, both factual and evaluative. However, it is important to emphasize that the experience that is in play in evaluating, forming, modifying and rejecting evaluative beliefs, the experience that may convince us that a certain evaluative view is true or false and lead us to a set of such beliefs that can truly be said to be our own, is denser, and broader, than in the case of non-evaluative beliefs about matters of fact. It is closer to the sort of experience of life that some old people are said to enjoy. The term ‘wisdom’ would seem to be more appropriate than ‘knowledge’ to designate the kind of intellectual achievement that this important cognitive faculty is directed at. We have insisted on the central importance that having a correct set of evaluative beliefs has for our happiness, flourishing and wellbeing. In this connection, we must equally emphasize, in connection with this, the importance of cultivating and developing the cognitive faculty we are referring to, for whether we end up having right evaluative views or not is strongly dependent on whether, and how, we exercise such a faculty.

Some remarks about this faculty whose aim we have called ‘wisdom’ may cast some light on the justification of our evaluative beliefs, as well as on the way in which the problem of regress might be avoided in connection with them. Think of a basic true evaluative belief, namely that it is wrong to cause unnecessary suffering to innocent people, and compare it with a basic theoretical belief, such as that two plus three equals five. Though in both cases a sort of immediate seeing or direct intuition (in something like the way Descartes uses this term) is involved, there are some important differences in the quality of the intuition. Part of the difference has to do with the role that emotional states play in each case. The way we ‘see’ that the evaluative practical belief is correct has to do with our imagining or thinking of actual cases that contravene it and the strong repugnance that they arouse in us. Unlike the case of the theoretical belief, our rejection of such contravening cases is not appropriately conveyed in connection with the term ‘absurd’, but rather with the term ‘abject’. It is, so to speak, a whole bodily rejection, not totally unlike our reaction to a putrid, ill-smelling piece of meat. We suggest that it is this holistic emotional involvement in our intuition of evaluative facts that actually prevents a regress of justification from arising in the case of evaluative beliefs. Concerning the evaluative belief we are referring to, someone's demand for a further justification could rightly be taken to reveal a cognitive and emotional impairment. To accept such a belief as simply true may actually be a requisite for someone to qualify as an agent whose evaluative moral views could be taken seriously, and so as a potentially morally responsible agent. In a different though related sense, it might be argued, as Christopher Hookway has actually done (cf. Hookway 2003), that affective states also play a central anti-sceptical, regress-stopping role in the case of factual and theoretical beliefs.

There are people who cannot actually see that certain basic evaluative propositions, such as the one that constitutes the object of the belief referred to, are simply true. They lack the faculty that leads to what we have called 'wisdom' and are not able to reach justified evaluative beliefs. Though different evaluative beliefs may be true, and the assessment of truth and falsity in this field is bound to be complex, evaluative beliefs that are obviously false may call into question the quality of their holders as moral agents. A considered judgement about this matter, however, should also take into account whether an agent could have reached alternative, true beliefs. This will be the theme of the next section. However, we may now say that a severe lack of wisdom, in the sense indicated, excludes such alternative possibilities and therefore deprives the agent of the control that would be needed for her to count as a morally responsible agent.

Beyond these minimal thresholds, the importance of developing and perfecting wisdom has to do with the fact that there are many evaluative beliefs whose truth is much less obvious than the one we have considered, as well as with the fact that the application of even obviously true evaluative views to particular cases is not equally obvious itself. In some cases, it may be difficult to ascertain whether a particular situation is or is not actually a case of causing unnecessary suffering to innocent people, to use the same example. We shall come back to this point in the fifth section.

Let us finally address the fourth condition, namely that evaluative beliefs should be potentially efficacious in our behaviour. We have pointed out above that, if our cognitive proposal about the possibility of moral responsibility for our actions is to work, it is important that our evaluative beliefs be able to motivate us and give rise to effective desires to act in accordance to them. If we propose to ground ultimate control over our choices and actions in our evaluative beliefs, they should be able to motivate such choices and actions. Otherwise, even if we controlled our evaluative beliefs and were responsible for them, we could not thereby be responsible for our choices and actions. Now it is a central tenet of a Humean view of motivation that only desires, and not beliefs, are able to motivate. We have argued that, if this view is correct, ultimate control, and so moral responsibility in the sense of true desert, will be undermined, and we criticized Kane’s apparent acceptance of a Humean view as jeopardizing his own libertarian perspective. As we pointed out with respect to the problem of justification, the question whether beliefs of an evaluative sort are able to motivate is a general issue with large ramifications, which extend far beyond the subject of this essay. However, we suggest that part of an affirmative answer to the question has much to do with the sort of cognitive access to evaluative facts that we have been talking about in the preceding paragraphs. Although there is discussion about the motivational potential of beliefs, it is largely agreed that emotional states can be motivationally effective. Now, it is *partly* because, and in so far as, we assess, correct and shape our evaluative beliefs through a denser and broader sort of experience than that related to matters of fact, one which involves our emotions, that the resulting beliefs can motivate us and give rise to desires to act in accordance with them. It is *partly* because, and in so far as, we judge that a certain evaluative belief is true or false on the basis of an experience that deeply involves our whole self, including our emotional system, that the resulting belief can have effects on our decisions and actions. Too narrow a view of both belief and emotion might be at the root of the widely shared Humean view that beliefs are not, as such, able to motivate an agent's decisions and actions. However, a complete treatment of this question would have to be the subject of a different essay.

Let us go on to comment on the fifth condition, that is, the condition of alternative possibilities or regulative control, whose necessity for moral responsibility we have defended against several lines of attack.

EVALUATIVE BELIEFS AND ALTERNATIVE POSSIBILITIES

We have extensively defended – especially in Chapter 2 – an alternative possibilities requirement for moral responsibility against several attacks on it. We argued that, concerning moral responsibility for actions, we rightly take into account whether the agent had a choice, whether it was in her power to decide and to act otherwise. We have insisted that this requirement is related to the control we want to have over the moral responsibility we bear for what we do. Unless we have alternatives, we lose such control, for we may find ourselves morally responsible for something that we could not have avoided doing or bringing about. In the context of our cognitive proposal, the representation of alternative possibilities of action takes quite a standard form. Our evaluative beliefs form a complex system; they comprise more than our moral beliefs. We also think, for example, that our self-interest and wellbeing is worth promoting; and we sometimes face situations in which different evaluative beliefs, all of them our own, point to opposite directions and cannot be jointly honoured. Moreover, we also happen to have desires and urges that we may think are not worth promoting and acting on, but we may sometimes indulge in acting on them. This projects a rather classic picture of what having alternative possibilities of decision and action amounts to in connection with moral responsibility.

But if we also think that a deep, ultimate form of control over our actions is required for our moral responsibility, it seems that alternative possibilities should extend, as a component of such control, to the roots or springs of such decisions and actions as well. In the cognitive perspective we favour, moral responsibility is grounded in a subject's evaluative beliefs. We should try to show that we rightly take into account whether an agent could have had different evaluative beliefs in order to judge whether she is morally responsible, morally praise- or blameworthy, for having the beliefs that she actually has and, derivatively, for acting in the light of them.

However, we have rejected epistemic voluntarism, according to which we can have a belief just by choosing or willing to have it. In fact, part of the motivation of our proposal is to see whether scepticism about moral responsibility could be avoided, and we have suggested that such scepticism may arise out of a conative or volitive perspective on moral responsibility, as centrally grounded in choices or acts of will. So, as applied to our evaluative beliefs, the requirement of alternative possibilities should not be understood in terms of choice. We should not require that, in order to be morally responsible for her evaluative views, and for the actions that she performs in the light of them, a subject could have *chosen* to have different beliefs. This would be at odds both with our rejection of doxastic voluntarism and with the spirit of our proposal. In this context, we seem to have two options. We can hold that the alternative possibilities condition applies directly only to our decisions and actions, especially to those decisions and actions related to our cognitive activity, and only indirectly and derivatively to our evaluative beliefs. Or we can hold that the requirement applies directly to our evaluative beliefs as well, and see what it means to say, in either case, that an agent could have had different evaluative beliefs. We can certainly interpret this locution in terms of the first option. On this interpretation, what we mean by it is that the subject could have carried out her cognitive activity in a different way, for instance by collecting more (or less) evidence or assessing more (or less) carefully the evidence she actually had. I think that this is what we sometimes mean by this locution. The question, however, is whether this is the only thing we mean or even could coherently mean by it. If the answer is affirmative, this gives the will a central role in the foundations of moral responsibility, which in turn, as we have tried to argue, might have harmful consequences for the prospects of avoiding scepticism. Taking this option, then, does not seem satisfactory. But I think that the locution can have a different meaning. We sometimes blame or praise someone for her moral views, and not just for the cognitive activity through which she reached them, which we may not know anything about. And in this case it seems that we are assuming that the subject could have had different moral views as a basis of our praise- or blameworthiness ascription. As a first clue to what that meaning may be, let us remember the importance we gave, for what concerns a non-volitive form of control over beliefs, to the attitude of respect and humility towards the real nature of the thing itself that one wants to have true beliefs about. But let us try to get a more intuitive sense of what that meaning may be by considering some examples where the question whether an agent could have had different evaluative beliefs plays a central role in the moral responsibility she bears.

Think first of a landowner in Ancient Greece, 4th century BC, who buys a slave to work on his property. Suppose that the landowner, call him Meno, is psychologically and cognitively normal. Let us also assume that buying slaves is a morally wrong action. On these assumptions, is Meno morally blameworthy for doing so?

Let us see what the judgement about this case would be from the perspectives of compatibilism and our own proposal.

Think of compatibilism. Suppose that determinism is true. It is plausible to assume that the condition of control, on a compatibilist construal of it, is met by Meno: he bought the slave because he decided to, and his decision was made for appropriate reasons. He also satisfies a conditional, compatibilist construal of the alternative possibilities condition: if he had decided not to buy the slave, he would not have bought him. He was also appropriately responsive to reasons (Fischer): if he had faced important reasons for not buying the slave, he would have recognized those reasons and would have decided and acted accordingly. It seems, then, that, on the correct assumption that buying slaves is morally wrong, the verdict of compatibilism should be that Meno was morally responsible, morally blameworthy in fact, for buying the slave.

From our cognitive perspective, however, the verdict about this case differs. Meno, the landowner, did something morally wrong in buying the slave, but he is not morally blameworthy for doing so. We may assume that he was the author of his evaluative beliefs, which included the belief that buying slaves is not morally wrong. We may also assume that he had alternative possibilities of decision and action, in a conditional and even in a categorical sense. But he did not have relevant alternative possibilities concerning his evaluative belief that buying slaves is not morally wrong. His belief was false and, in acting on it, he was doing something morally wrong. But he was not morally blameworthy for his action, because, given his historical and social circumstances, the true belief that buying slaves is morally wrong was not within his possible cognitive landscape; it was not available to him. Even if Meno had formed his evaluative beliefs with the right attitude of humility and respect for evaluative facts, it is not reasonable to expect that he could have seen that truth, which started to emerge and spread only four centuries later, with Christianity. Even though Meno can be said to have control over his evaluative belief that buying slaves is morally permissible, and to be its true author, as these concepts were sketched out in the preceding sections, he did not have ultimate *regulative* control over that belief: *he could not have had a relevantly different belief*, namely the belief that buying slaves is morally wrong. And this is why the judgement should be that Meno is not morally blameworthy for buying the slave.

But I think that this is also the judgement that our natural, pre-theoretical attitude toward this case will yield.

A consequence of our perspective that can be seen in the discussion of this example is that, even if moral objectivism is true, so that the truth of moral propositions is not relative to particular circumstances and times, moral responsibility, instead, is relative to circumstances and times. To see this, think now of another character, Timothy, a landowner in South Carolina, mid 19th century AC, who also buys a slave who can work on his property. Assume, as before, that buying slaves is morally wrong and that Timothy was also psychologically and cognitively normal. Suppose also that he thinks, like his Greek counterpart, that buying slaves is not morally wrong. Timothy thinks so, we may suppose, owing to his upbringing in a slave state and environment. Now, though an accurate judgement about his moral responsibility would require filling in more details, I think that, with the elements at hand, the correct verdict in this case is that he is morally blameworthy for buying the slave. He does not think he is doing something morally wrong, but in this case he could, and should, have had a different belief. The truth that slavery is a morally wrong institution, and so that buying a slave is not morally permissible, was within his possible cognitive landscape. He, unlike Meno, his Greek counterpart, could have believed that truth. For Meno, this moral truth lay four centuries ahead, but for Timothy it was eighteen centuries behind. Moreover, he lived in a Christian environment, in which such a truth could be found relatively easily. And, for an educated man, the arguments against it, concerning for example the inferiority of black people, could easily be seen as fallacious. Therefore Timothy is morally blameworthy for buying the slave, given that this action comes from a belief over which he had ultimate regulative control.

In the light of all this, let us see what having direct regulative control over one's beliefs may consist in. In saying that Timothy, the American landowner, could have had a different evaluative belief we are not simply saying that he could have carried out an appropriate and more careful cognitive activity, though we may also mean this. We are also saying that he could (and ought to) have *seen* what was there to be seen. He was not humble and respectful enough towards patent evaluative facts. There are some connections between this judgement and cases in which we blame someone for something that she involuntarily did or omitted. Our partner may blame us for forgetting her birthday or for forgetting an appointment for dinner we had agreed. Forgetting is obviously not voluntary. It is not even an action. Our omission is not voluntary either. But I think this ascription of blame is not like blaming someone for, say, being ugly. The latter judgement is not fair, but it seems to me that the former can be. And what explains the difference is the assumption that we have a form of control over our forgetting that we do not have over our physical constitution. In some sense of the words, it is true that we ought not to have forgotten our partner's birthday or the appointment we had for dinner, and it is also true, in some sense of the words, that we could have remembered both things. When we say that we could have remembered such things, we are not merely saying that it was logically possible that we should remember them. And when we say that we ought to have remembered them, we are speaking of moral obligation. These 'could' and 'ought' are those that feature in the 'ought'-implies-'can' principle and in the alternative possibilities condition of moral responsibility. We are saying that remembering those things was our duty and that it was within our reach or power. And the regulative control that is being assumed in the corresponding blame ascriptions is, I think, close to the regulative control we assume people sometimes have over their beliefs, including their evaluative beliefs. It is the sort of non-volitive regulative control that we have been suggesting.

Even if believing, like forgetting, is not voluntary, and not an action, we can justifiably blame someone for her beliefs, as our partner can justifiably blame us for forgetting our appointment. The analogy goes quite far, which suggests that we are facing closely related phenomena. In blaming us for forgetting our appointment, our partner is blaming us for our blindness to both our appointment and her, which reveals our lack of respect for both. Similarly, in blaming the American landowner for his belief that slavery was morally permissible, we are blaming him for *not seeing* what was there to be seen, namely the moral evil in slavery and the slaves themselves as his fellow human beings, which in turn reveals his lack of respect for them.

If we agree that it makes sense and can be justified to ascribe moral blame to someone for forgetting an appointment (and not, say, for her ugliness) and to assume, as a basis for that judgement, that she ought to, and could, have remembered it, then we are acknowledging that there is a form of regulative control that does not depend on choice and that is relevant to moral responsibility. It is plausible to think that we have this sort of control over our beliefs and that we assume that persons have it when we morally blame (or praise) them for their moral and evaluative beliefs, and not only for their actions. It would be important to go from the rough idea of this sort of control which we have derived from the examples to an investigation of its precise structure and conditions. But, on the basis of the preceding examples, we have reason to think that it exists. And, if our diagnosis of the roots of scepticism about moral responsibility is correct, its existence may be significant for the prospects of overcoming the sceptical predicament.

Seeing something that is there is not an action, but it is something for which one can sometimes deserve merit, as is the exercise of a capacity that can be cultivated and perfected. This is especially true with respect to evaluative facts. Respect for others is a basic moral insight that most people get from their social environment, and it provides a reason for developing that capacity and improve our receptiveness and sensitivity toward the needs and value of other human beings. Many circumstances can and should be taken into account in evaluating an agent's praise- or blameworthiness for seeing, or not seeing, certain evaluative facts, and for having the corresponding beliefs. More favourable circumstances may lead to a more rigorous judgement about an agent if she does not develop right evaluative views, and less favourable circumstances may assuage or even preclude an agent's blame for not doing so. Conversely, an agent who develops her sensitivity to evaluative facts in a difficult environment may deserve more praise than someone raised in a more hospitable situation.

Though our examples have dealt with blameworthiness, alternative possibilities are also relevant to an agent's praiseworthiness for her evaluative views. This assertion goes against Wolf's asymmetry thesis, according to which alternative possibilities are required for an agent's blameworthiness, but not for her praiseworthiness. In Chapter 2, we argued for the general application of the alternative possibilities requirement with respect to actions. And we think that something similar can be said about evaluative beliefs. Just as an agent deserves blame for her wrong evaluative views in so far as it was in her power to have formed different, and better, views, an agent deserves praise for her correct evaluative views in so far as she could have formed a worse set of such views. These judgements should be qualified on account of several factors, including the circumstances in which she grew up and her natural gifts, but it seems to me that the consideration of those alternative possibilities informs and underlies such judgements.

It may seem that, in requiring the availability of alternatives for an agent's praiseworthiness for her actions and evaluative beliefs, we are assigning a positive value to a kind of freedom, namely the freedom to act wrongly or to form mistaken moral beliefs, which, in Dennett's expression, would not be 'worth wanting'. But I think this compatibilist move is wrong. In requiring alternatives in order for an agent to be praiseworthy for her good actions or correct evaluative views, we are not giving a positive value to the possibility of acting in a bad way or forming wrong evaluative views. We give a negative assessment of this possibility, but, at least for human beings, the possibility is there, and it is in contrast with it that we praise an agent for not taking it. In a related vein, Wolf (1990: 81–2) rejects a consequence of such a requirement, namely that a person whose moral convictions and commitments are so strong that she is literally incapable of acting in a morally wrong way is not praiseworthy for her good actions. This consequence, however, is not obviously false. For us human beings, there is pressure, coming from some of our desires and natural inclinations, towards morally wrong ways of acting and believing. In the absence of such a pressure, in the absence of any countervailing factors to our invariable disposition to form objectively correct moral beliefs and act on them, in the absence of any temptation, we would be good persons but we might not actually deserve praise for those beliefs and actions. Just as a person can perform a morally bad action and not be blameworthy for doing so (remember our Greek landowner), a person can perform a morally good action and not be praiseworthy for doing so either. And the reason, in both cases, is that they lacked regulative control: they could not have believed or acted otherwise. This is partly why we judge that subjects in a Platonic New World, who are determined to have right values and to act on them, are not morally praiseworthy, even if they are good people. True authorship, as a basis for true desert, requires the availability of alternatives.

Whether a person has deep regulative control over her choices and actions, and so is morally responsible for them, is difficult to establish. This means that our moral responsibility ascriptions may be wrong in various particular cases. But in focusing on evaluative beliefs, rather than choices and willings, as the root of moral responsibility, we are trying to show that moral responsibility is indeed possible, even if its presence and degree may be hard to assess in many or even most cases. A crucial test for our incompatibilist, libertarian proposal is whether it can meet the classical compatibilist objection against libertarianism, according to which indeterminism undermines the control that would be required for moral responsibility. Remember that meeting this objection is the sixth and last condition we have required for the plausibility and anti-sceptical effectiveness of our cognitive view of moral responsibility. The ‘Mind’ argument, in its different versions, is the standard formulation of this objection. Let us now address this complex and important question.

INDETERMINISM, BELIEF AND MORAL RESPONSIBILITY

The purpose of this section is to see whether indeterminism can be reconciled with rational control. We shall distinguish two perspectives on the place and role of indeterminism in practical rationality and contend that, though one of them is unlikely to provide a satisfactory response to worries about rational control in indeterministic contexts, the other might well be able to do so. On this basis, we shall be arguing that a cognitive approach to moral responsibility has better chances than a conative, will-centred approach of giving a justified positive answer to that question.

Many libertarians see a metaphysical basis of free will and moral responsibility in quantum indeterminism, especially as it may take place in the human brain. Kane, as we have seen, is among them. If quantum indeterminism at the subatomic level were amplified, instead of cancelled, at higher levels of the organization of matter, such as the neuronal level, this would make the brain into an indeterministic system, so allowing for alternative neural pathways and, on appropriate views of the relationship between brain and mind, for the alternative choices and actions that, according to libertarians, including the author of this essay, are required for moral responsibility.

Now, though we accept the alternative possibilities requirement, we do not share those libertarians' hopes about the perspective we are considering. On this perspective, indeterminism at the subatomic level would be amplified and transmitted to higher neurological and mental processes through metaphysical relations such as supervenience or even identity. We can call this view of the role of indeterminism in mental acts and processes ‘bottom-up’ indeterminism. We do not see how to avoid the conclusion that, even if it provides room for alternative pathways of decision and action, bottom-up indeterminism sweeps away the agent’s rational and volitional control over them, so eroding the basis of moral responsibility. Quantum phenomena are among the most basic physical processes. As we pointed out in the preceding chapter, if quantum changes suffered by a subatomic particle are actually indeterministic, they are not under anyones’s control. So any attempt to ground free will and moral responsibility in such basic indeterministic processes will stumble, at some point, on the problem of the agent’s control over her choices and actions. It would seem, then, that the versions of libertarianism that accept bottom-up indeterminism as a metaphysical basis for free will and moral responsibility are bound to fall prey to some construal or other of the ‘Mind’ argument, whose last consequence is SMR’s premise C, namely the impossibility of moral responsibility given indeterminism. The fact that a quantum-cum-chaotic process occurs in my brain which is identical or subvenient to a certain decision does not make that process and decision truly my own, in the sense that would be needed for moral responsibility, and neither does it allow me to rationally control them. Both aspects of ultimate control, namely ultimacy of source and rational control, are actually undermined by bottom-up indeterminism.

If indeterminism is to make free will and moral responsibility possible, it has to hold at the right places and have the right relations to the mental and neurological phenomena, so as to provide room for rational control. We propose to call an indeterminism of this sort ‘top-down’ indeterminism. According to this perspective, the sort of indeterminism that can allow for rational control holds primarily at ‘higher’ levels of reality, especially at the level of rule-governed, normative systems, practices and institutions, such as language itself, as well as written or unwritten codes regulating various aspects of social interaction within human societies. These rule-governed practices include practical deliberation. What has to be shown is that, at this level, indeterminism holds and is nonetheless compatible with rational control over alternative pathways. As a second step, we should tell a reasonable story about how indeterminism at this higher level could be transmitted downwards, so as to contribute to the shaping of the human brain and some of the neurological processes that take place therein. As the reader will have noticed, this second step involves an attempt to deal with the problem of mental causation, which is an aspect (the possibility of mind-to-body causal influence) of the traditional mind-body problem. Our aim concerning this venerable and recalcitrant problem will be quite modest. We shall try to show that downward causal influence, from mental, content-bearing states or events to physical ones, is neither unintelligible nor necessarily at odds with a scientific outlook.

Let us go on to the first task. One sense in which normative systems, practices and institutions can be said to be indeterministic is that, as Peter Winch (1963), following Wittgenstein, insisted, the distinction between correct and incorrect ways of acting in relation to the corresponding normative standards is constitutive of them. But this distinction involves the notion of alternative possibilities. They can be said to be indeterministic in a second and important sense, namely that, in several particular cases, the system's normative standards do not determine one single way of complying with (or breaking) them and of acting in a correct (or incorrect) way; in relation to this, in many cases there is wide room for reasoned argument and discussion about whether a particular way of acting is or is not correct with respect to the system’s normative demands. This is why legal codes, for example, essentially need specific institutions to apply them to particular cases. This is not due to the fact that the codes are not sufficiently precise; it is rather within the nature of a rule or norm that the question of what it means to comply with it can always arise in certain cases. So, in the administration of justice, reasoned verdicts about particular cases (jurisprudence) become an integral part of the normative criteria for further judgements. In relation to this, the question whether a particular act is or is not justified with respect to certain normative criteria can sensibly be raised on many occasions. The concept of justification is, then, essentially open-ended and relative to circumstances of several kinds. Finally, normative systems and practices are indeterministic in the sense that there is room for reasoned discussion about whether a particular situation is one to which a certain normative judgement applies. To take a previous example, even if one accepts the principle that it is morally wrong to cause unnecessary suffering to innocent people, this does not decide the question whether a particular situation actually falls within that principle’s scope.

The significance that this reference to normative systems has for our proposal about the possibility of moral responsibility partly has to do with the fact that our evaluative (and non-evaluative) beliefs also form a system that is subject to normative standards related to the constitutive aim of truth. Demands for coherence or responsiveness to the relevant evidence bear upon human beings as subjects of beliefs. Our belief systems are indeterministic in the three senses we have distinguished in the preceding paragraph. The distinction between correct and incorrect, justified and unjustified beliefs and ways of forming them is constitutive to the possession of such a system. There is also room for discussion about whether and when a certain belief or belief set complies with the normative demands and so is actually correct or justified. This second feature, according to which, in many particular cases, the normative standards do not determine what it is to comply with them and so whether a certain belief is correct or justified, makes it the case that the holding of particular beliefs in certain paradigmatic or extreme cases often acts as a negative touchstone of the ability for such a compliance. As we saw in the preceding section, rejection of certain basic evaluative beliefs may disqualify a person as a morally responsible agent, for we take this rejection to be a clear sign of her inability to conform to normative standards and form correct moral beliefs. Our previous example of someone's seriously holding that there is nothing morally wrong in causing unnecessary suffering to innocent people may be a case of this kind. Finally, it also seems clear that there is room for reasoned discussion about whether a certain belief is or is not a case to which a doxastic normative standard applies.

It is not only our system of evaluative beliefs that is subject to normative standards. Our practical deliberation, by which we try to evaluate possible ways of acting and reach decisions in the light of those assessments, is a practice that is also subject to normative demands. As Christopher Hookway writes, 'reflection, including both practical and theoretical deliberation, is an activity that can be controlled through the exercise of normative standards' (Hookway 2001: 190). Central among those standards are our evaluative beliefs about the worthiness of several ends and purposes. And just as, concerning our evaluative beliefs, there is room for reasoned discussion about whether a particular belief does comply with the normative standards of coherence or responsiveness to facts and experience, or about whether a particular belief falls within the scope of a standard, there is a similar free space concerning our practical judgements and decisions for what concerns their relationship to our evaluative beliefs. For example, for an agent who highly values friendship it is still an open question whether a certain practical judgement or a decision about a particular way of acting is actually one that honours that belief or is the one that best honours it. And it may also be an open question whether a certain way of acting violates that evaluative belief.

Thus we have at least two spheres closely related to moral responsibility (our evaluative beliefs and our practical deliberation) to which normative standards constitutively apply, and each of them shows at least the three forms of indeterminism that we have indicated concerning normative systems and practices in general. Moreover, in spite of the hierarchical ordering between these two spheres, practical deliberation can also retrospectively affect our evaluative beliefs. For example, a decision against one of our evaluative beliefs, perhaps prompted by a momentary impulse or special emotional state, even if it is presently experienced as wrong and accompanied by a feeling of guilt, may end up showing us that it was the belief, not the decision or the corresponding action, that was actually incorrect. And, given the inner dependence relations among our evaluative beliefs, this can lead to a reorganization of our whole system of such beliefs or a part thereof.

The complex net of interactions between these two normatively guided spheres, each with its own field of alternative possibilities, opens up an even wider indeterministic field of free movement for us as believers and practical deliberators. However, it is centrally important to realize that this indeterminism related to normative systems and practices, unlike the indeterminism related to quantum phenomena, is essentially linked to reasons and reasoned discussion and criticism, and thereby initially receptive to rational control. It is the sort of indeterminism that, being essentially associated with the possibility of rational argument, may give rise to the sort of non-arbitrary, reason-related freedom that can ground moral responsibility. This is the sort of freedom that grows in what John McDowell, following Sellars, has aptly called 'the space of reasons'. As he writes, 'this freedom, exemplified in responsible acts of judging, is essentially a matter of being answerable to criticism in the light of rationally relevant considerations' (McDowell 1998: 434).

This wide field of possibilities of belief, judgement and decision, backed by reasons, also opens up the space for forging rationalizations about the true motives of our choices and actions, inventing excuses for what may appear as morally unjustified and blameworthy actions, and even for self-deception about our actual reasons or motives. These rationally flawed phenomena pay retrospective tribute to reason, however, in that they are only possible and acceptable in so far as they present themselves in the guise of true and rationally justified explanations.

Moreover, normative systems are significant to the possibility of moral responsibility, not only in connection with the rationality aspect of ultimate control but also with regard to its ultimacy of source aspect. We have argued that, in appropriate conditions, we can truly be considered as authors of our evaluative beliefs, even if they trace back to factors external to ourselves. We may now see how external factors can actually be indispensable to our status as subjects of evaluative beliefs, if responsiveness to normative standards is constitutive of such a status. For it is only through socialization and education in a human society that we can become participants in normative systems and practices. And it is as participants in such systems and practices that the idea of alternative possibilities and of reasons for each of them, as well as the idea of rational justification and control over our beliefs and choices, make sense to us. It is as social beings, as members of objective systems governed by normative standards that we become capable of rational control, deliberation and reflection. And, as these abilities are a condition of someone's qualifying as a morally responsible agent, it is as social beings that we can be agents to whom moral responsibility ascriptions can justifiably apply. As we have insisted in previous sections, the social nature of human beings should be viewed as an enabling condition for their moral responsibility rather than as an obstacle to it. We can now give further support to that contention.

Let us now move on to the second step in the analysis of top-down indeterminism as an enabling ground of moral responsibility. Even if normative systems and practices can rightly be taken to be, in McDowell's words, 'second nature' to us, it is important to see whether the reason-related indeterminism that we have found in them could reasonably be taken to transmit itself downwards, so as to partially shape our 'first nature' as natural biological systems. Otherwise, if reason-insensitive determinism or quantum indeterminism reigns unopposed at the neurological level, unaffected by the reason-sensitive freedom of the 'space of reasons', our proposal might fall prey to the objections of scientific obscurantism traditionally levelled against libertarianism, or be open to the charge of erecting a problematic form of dualism, on the basis of an unrestricted opposition between freedom and nature. And these objections, in turn, might prompt the sceptical suspicion that belief in freedom and moral responsibility, psychologically inevitable as it may be, is nonetheless an illusion.

Let us think of some examples of normative systems and practices, such as musical notation and its interpretation with a musical instrument, or arithmetic and its application by a shop assistant to calculate the amount owed by customers. In both cases there is a body of theory with an objective content and rules. And anyone wanting to master those systems should be prepared to learn that content and be trained in applying the rules. Both cases show, in different degrees, the levels of indeterminacy we distinguished when we referred to normative systems in general. Now think of learning to play a musical instrument, say the classical guitar, after having being trained in reading musical notation. One has to translate musical signs into sounds by pressing the strings with one's fingertips on the right places of the guitar neck while simultaneously plucking the right strings with the fingertips and nails of the other hand. In the first stages of this learning process, the movements are clumsy and difficult to perform; one has to think of the steps involved in interpreting each note. Gradually, however, through a painful process and a good deal of practice, the movements of the fingers become less reflective and more automatic; the reading and translating of the musical notation into sounds is less of a problem, so that one can then concentrate on other aspects of the interpretation, such as expressiveness or the clarity of the sound; aesthetic pleasure gradually increases and playing becomes more rewarding. As one confronts a difficult new piece, the finger movements are not as sure as in pieces one has already studied, but, also gradually, the process of managing to play it decently becomes shorter and less painful than on prior occasions.

Let us stop here in this (not too inexact, I hope) phenomenological description of the process of learning to play the guitar. Now the question is what happens during this learning process at the neurological level. Here we must leave the firmer ground of description and proceed to the shaky terrain of speculation. But we do not aspire to empirical accuracy. It will be enough if we are able to suggest a reasonable and fairly general hypothesis not at odds with plausible neuroscience. Here it is. Corresponding to the progress in the learning, what takes place in the brain and the nervous system is the establishment of an increasingly complex system of neural connections, which include afferent and efferent nerve fibres. These connections are reinforced with practice, which explains the increasing sureness and automatic character of one's movements during the process of learning, and they are also correspondingly weakened by lack or insufficiency of that practice, which in turn explains why, after a long period with no playing, the performance becomes poorer and the movements clumsier. If the period is very long, the subject can even find herself no longer able to play pieces she used to play quite easily in the past. In the end, she may find that she needs to go through the learning process again almost from the start. This second process, however, is likely to be shorter than the first, owing to the remaining neural connections. In a nutshell, at the neurological level the process of learning consists in the forming of a certain configuration of neural networks.

We can now go back to philosophy. It is the neural configuration that, at least partially, explains someone's ability to play the guitar – which is a particular normative practice – with a certain degree of decency. However, what at least partially explains the fact that this neural configuration, with its various alternative pathways and complexities, has been formed, as well as the form it actually has, is the objective content of normative systems, institutions and practices related to music and guitar playing. Through the process of learning, the several indeterministic levels involved in those systems and practices have come to shape the structure of the brain and nervous system of the learner. New forked neural pathways have been shaped through learning, according to the alternative pathways corresponding to the levels of indeterminism of the normative systems and practices involved in the learning. And, since the learning never actually ends, this indeterministic shaping of the brain goes on with it, becoming more complex and subtle, and more sensitive to tiny variations in neural activation levels. However, since this branching neural structure has been shaped, through the learning process, according to the reason-sensitive alternative possibilities constitutive of the normative systems and practices involved, it does not undermine the agent's rational control over those alternatives, but rather enables and enhances it. At the same time, however, some physical and biochemical processes in the brain unrelated to reasons, such as natural decay or death of neuronal tissues, hormonal insufficiency or even indeterministic changes at the quantum level, act in the opposite direction, increasing entropy and damaging rational control.

Note that this (admittedly rough) picture of the relationships between the 'space of reasons' and the brain, between mind and body, to put it in more traditional terms, is compatible with supervenience of mental properties on physical ones. Think of an agent in a certain phase of her process of learning to play the guitar and imagine that an exact physical duplicate of her suddenly comes or is brought into existence. Given supervenience, this creature will also be able to play the guitar with the same level of perfection as the original agent. But some comments are in order. First, I take it that the probabilities of such a physical duplicate coming to exist without herself going through a process of learning are astronomically low. Moreover, unless the new creature engages in a process of learning to play the guitar, so that her inborn abilities are cultivated and perfected, she will quickly cease to be a physical-cum-mental duplicate of the original agent. Finally, though the original agent is truly praiseworthy for her ability to the play guitar with a certain degree of perfection, her physical duplicate is not. This, however, does not deny supervenience. It is rather the case that some properties, both mental and physical, are historical and dependent on their origin and the actual causal interactions that gave rise to them, so that they supervene on a broader physical basis. Being a sunburn, to take a Davidsonian example, is one of them. An alteration of the skin physically identical to a real sunburn but not caused by the sun is not itself a sunburn, because it does not have the right origin. Being a planet, a Fodorian example, is another. As for mental properties, some of them also show this feature. Remembering something is one example. Being praise- or blameworthy (including moral praise and blame) for a certain action or achievement is another. Now, if the physical duplicate of our musical heroine had come to her present physical state, which enables her to play the guitar, through the same physical history and causal interactions as the original agent, she would also be praiseworthy for her present ability and performances, for she would also have gone through a learning process and those achievements would rightly be attributable to her. If, however, she had been suddenly brought into existence, as we initially assumed, she would not deserve praise for those abilities and performances. This, incidentally, is at least part of the reason why Brave New World citizens, provided that they have come to have their evaluative views through brain manipulation, are not morally responsible, praise- or blameworthy, agents.

With due respect for the differences, the preceding remarks can also be applied to many other normative systems and practices, including those directly related to moral responsibility, such as an agent's system of evaluative beliefs and her practical deliberation.

Now, if top-down indeterminism, as we have characterized it, actually holds, that is, if, as we have argued, normative systems and practices constitutively incorporate alternative pathways backed by reasons at the different levels we distinguished, and if this reason-sensitive indeterminism can actually transmit itself downwards so as to partially shape the inner structure of our brains, then, provided that our evaluative beliefs and practical deliberation are normative in that sense, we can have the makings of a response to the traditional compatibilist objection to incompatibilism, namely that indeterminism precludes rational control over our decisions and actions and, with it, moral responsibility itself. Let us see how this worry might be dispelled.

We shall contend that the objection is powerful, perhaps decisive, against traditional, will- or choice-centred versions of libertarianism, especially when combined with what we have called a bottom-up conception of indeterminism, but that a cognitive libertarian approach of the kind we have suggested might well be able to meet it. We shall address this worry (the ‘Mind’ argument) with the aid of several versions of it that we went through in the first section of Chapter 4.

Remember Van Inwagen’s example of the judge (J) who, after calm and careful deliberation, decides not to raise his hand at time T and so not to grant special clemency to a convict. J does not raise his hand at T and the convict is sentenced to death. As we have seen, we are indebted to Mele for one powerful way of formulating the ‘Mind’ argument on the basis of examples of this kind. Adapting Mele’s suggestion to Van Inwagen’s example, imagine a close possible world, with the same past and natural laws as the world J inhabits, in which a twin of J’s, call him J*, exists. Suppose now that the only difference between these two worlds comes at time T. At that time, while J decides not to raise his hand, J* decides to raise his. Remember Mele’s words: ‘If ... there is nothing about the agents’ powers, capacities, states of mind, moral character, and the like that explains this difference in outcome, then the difference really is just a matter of luck’ (Mele 1999: 99).

We do not see how to avoid Mele’s conclusion if we hold, as libertarians traditionally do, that, for moral responsibility to be possible, a different choice has to be compatible with the same past and natural laws that led to the agent’s actual choice. This contention is a form of the categorical interpretation of the alternative possibilities condition on moral responsibility. To see what is wrong with this understanding of the condition, let us go back to the example of J. As Van Inwagen presents the example, J goes through ‘a suitable period of calm, rational, and relevant deliberation’ (Van Inwagen 1983: 69). So we can plausibly assume that, at a certain time in his deliberation, J arrives at a practical judgement according to which, all things considered, it is best to deny clemency to the convict, and so not to raise his hand, or at least better than the relevant alternative. Now, J* also arrives at this practical judgement and, unlike J, decides to raise his hand. Given that J and J* are perfect twins until the time of the choice, this presents J’s choice as purely chancy and arbitrary. He might equally have decided to raise his hand. So, if we insist that the choice is undetermined until the very moment it is actually made, we are severing the link between practical judgement and choice. We are accepting that, for moral responsibility to be possible, an agent must be able to decide to do B even though she judges that doing A is better. Decisions of this sort may happen if, for example, cases of weakness of the will do actually occur, but weakness of the will is, in everybody’s lights, an irrational phenomenon. To insist on this sort of movement, then, amounts to grounding moral responsibility in arbitrary or irrational choices and paving the way for the compatibilist objection. Indeterminism placed at this particular juncture, that is, between practical judgement and decision, certainly erodes the agent’s rational control over her choices. To resist Mele’s objection by holding that, in quantum-cum-chaotic scenarios, sameness of individuals or their stories is not defined, as Kane does, actually worsens things, for that move strongly suggests that a bottom-up conception of indeterminism is being assumed, and with it the dependence of choices on quantum changes in the brain, beyond the agent’s rational or even volitional control.

At least as powerful as this version of the objection is the ‘Rolling-Back’ version. In Van Inwagen’s presentation it features an agent, Alice, who, in a difficult situation, confronts a choice between lying and telling the truth, and freely chooses to tell the truth. Imagine that God reverts the world to the state it was in before that choice and allows it to go ‘forward’ again. Supposing that her act is undetermined, we can plausibly assume that, after a high number of replays, the ratio of each possible outcome with respect to the total number of replays converges on some value, say 0.5. On this basis, the only reasonable option is, for each new replay, to assign the same probability, 0.5, to Alice’s lying and to her telling the truth. So this was also the probability of Alice’s lying, and of her telling the truth, in the actual, initial situation. The conclusion, in Van Inwagen’s words, is that ‘an undetermined action is simply a matter of chance’ (Van Inwagen 2000: 15), as compatibilists have traditionally held. The argument works equally well for choices or decisions, instead of acts. The problem would seem to be roughly the same as in Mele’s version of the objection.

Now, how does our proposal fare with respect to these objections? Corresponding to the cognitive character of our approach to moral responsibility, and contrary to volitive approaches, the main role in the process of practical deliberation should be assigned to practical judgements rather than choices. In the picture we have sketched above, practical judgements about which action is best, or better than the alternatives, should be understood as the application of an agent’s evaluative beliefs, as normative standards, to a particular situation she faces. A practical judgement is a cognitive rather than a volitive phenomenon. It is not properly a choice, but, so to speak, the expression of a belief about the value of particular ways of acting, formed in the light of more general evaluative views. As a rule, an agent's rational choices are fixed by her practical judgements. Choices cannot ride free from practical judgements if they are not to fall into arbitrariness. On our cognitive approach, then, practical judgements replace choices as the central steps in practical deliberation. Choices, in turn, are dependent on practical judgements and remain linked to them. At a deeper level, which grounds the agent's ultimate control over her choices and actions, we have again, of course, beliefs – of an evaluative sort – rather than radical choices (e.g., Kane's self-forming willings). These beliefs are formed, in turn, in the light of another set of normative standards, such as coherence and sensitivity to the facts.

This picture is not a mere reorganization of the components of practical deliberation. Giving centre stage to practical judgements, conceived as applications of more general evaluative views to particular cases, allows retaining indeterminism in practical reasoning without giving up rational control. As we pointed out above, application of a normative standard to a particular case is not a mechanical task. It leaves space open for rational discussion about both whether a particular way of proceeding complies with the standard and whether the particular case at hand actually falls within the scope of the standard. It is, then, both an indeterministic process and a non-arbitrary, reason-sensitive one. In this context, we need not deny the libertarian view that choices are undetermined; but we view the indeterminacy of choices as derived from the reason-sensitive indeterminacy of practical judgements, so protecting choices themselves against charges of arbitrariness or irrationality.

With these considerations in place, let us return to the objections and see whether we can plausibly resist them.

Think first of Mele's 'twins' objection, which we applied to the judge example devised by Van Inwagen. We may assume, as Van Inwagen also does, that J is rationally and emotionally normal, and capable of competent practical reasoning. We may then assume that, at a certain moment in his 'calm, rational, and relevant' practical deliberation, he arrived at the practical judgement that sentencing the convict to death, by not raising his hand, was better than the relevant alternative. Assuming rationality again, we may also accept that, at a deeper level, he holds the belief that capital punishment is justified, at least for some especially serious crimes. In these circumstances, his decision not to raise his hand and his corresponding action are ones he has rational control over. (An additional question is whether he has ultimate regulative control over this belief, and so whether he is morally responsible for his practical judgement, choice and action. This depends on his having met the conditions of true authorship with respect to that belief, including the availability of the opposite – and, in my view, true – belief that capital punishment is not justified). Now imagine J's psychological twin, J*. The objection is supposed to be that, if choice is causally undetermined, J* may well make a different decision than J, namely to grant clemency by raising his hand, which shows that, contrary to the hypothesis, J's original choice was a matter of chance and so beyond his rational control.

Now, are we forced to accept this conclusion? In order to answer this question, we have to ask to what extent the sameness of both twins and their histories is supposed to hold.

If they and their respective histories were exactly alike until the very moment of choice, *including the practical judgement that not granting clemency is better than the opposite,* then the conclusion would seem to follow. This move, however, may work against some will-centred libertarian views, but it begs the question against our proposal, which considers practical judgement as the central step in practical deliberation, and rationally controlled choice as derived from it. Assuming exact sameness between the two agents until the very moment of choice, the difference between their respective decisions may be explained, in the context of our proposal, by pointing out that J*'s practical judgement, unlike J's, did not control his decision, which so was arbitrary and irrational. In fact, if J and J* form the same practical judgement, namely that not granting clemency, and so not raising their hands, is better than the opposite, and nonetheless J does not raise his hand while J* does, then J*, unlike J, acts against his best considered judgement and so in a weak-willed way. Being contrary to his best judgement, J*'s decision was irrational and arbitrary. But we cannot conclude that J's decision, just because it *might* also have been weak-willed, was equally irrational and arbitrary. Provided that J decided in accord with and owing to his practical judgement, his decision was rational and controlled by that judgement. That weakness of the will is always possible does not show that we never decide in a controlled way.

Even if an agent's practical judgement does not causally determine his decision, it is not mere luck that people's decisions usually accord with their practical judgements. When this is not the case, we need a causal/psychological explanation, which is not needed otherwise. The objection may implicitly assume that indeterminism undermines rational control by severing the causal link between practical judgement and decision. But the absence of causal determination does not erase the distinction between J's and J*s decisions. A way of expressing the difference is to say that J's decision, but not J*'s, was justified, and so non-deterministically caused,[3] by the relevant practical judgement. In terms of our prior distinction, we might say that J*'s decision was a matter of bottom-up indeterminism, while J's decision was made within top-down indeterminism.

So, if the sameness between the two agents is assumed to hold until the very moment of choice, the conclusion that J's decision is arbitrary and chancy does not follow. The objector can respond to this reply by accepting that J and J* form different practical judgements. He may then try to reinstate the objection by pointing out that, under the assumption of indeterminism, J and J*, reasoning *in exactly the same way,* can form different practical judgements, which reveals the ultimate arbitrariness of J's actual judgement. Leaving aside Buridan cases, if, in the light of the same standards and evidence, and through exactly the same process of reasoning, including their mentally spoken words, J and J* form different practical judgements, this would seem to show that J's actual practical judgement was a matter of chance. But our proposal suggests a different perspective on the situation. Assuming *exact* sameness between the two agents until the forming of their respective practical judgements, if the latter differ, it seems that at least one of these judgements (presumably J*'s) was not appropriately backed by the relevant reasons. J*'s practical judgement was, perhaps, a case of weakness of warrant, and so an irrational phenomenon in itself, independently of the question of indeterminism. Again, that J *might* also have formed his practical judgement through weakness of warrant, as J* did, does not show that J's actual practical judgement, provided that it was rightly controlled by his reasons (the reasons he shared with J*'s), was affected by arbitrariness or irrationality.

The objector intends to conclude, from the fact that J and J* are exactly alike, and that, if indeterminism holds, they can form different judgements or make different decisions, that, under the assumption of indeterminism, our judgements and decisions are a matter of chance, so that indeterminism is incompatible with rational control. But the conclusion does not follow, as we have tried to show. There is at least one alternative explanation, namely that J*'s practical judgement and/or decision, but not J's, were formed irrationally. It should be possible to distinguish, in an indeterministic world, between rational and irrational beliefs and decisions.

So our perspective can successfully meet the 'Mind' objection if this is formulated in terms of exactly alike twins. In cases of less than perfect sameness, we can certainly accept that two almost alike agents can form different beliefs and decisions that can be equally rationally justified. The sameness that can be assumed to hold between the two subjects can go very far. It may extend to the holding of the same relevant evaluative beliefs, including of course the belief that capital punishment is justified, at least in some cases, as well as the same degree of competence in practical and theoretical reasoning. It may include many other psychological traits as well. With this degree of assumed sameness, we can accept that J and J* may well form different practical judgements and make different choices. But this does not force upon us the conclusion that the difference, and with it J's original decision, is just a matter of chance, even if indeterminism holds. For, if we are right that application of normative standards to particular cases leaves room for reasoned alternatives, J and J* may differ, for example, in the question whether the convict's crime falls within the scope of their common evaluative belief, that is, whether it actually is serious enough to deserve capital punishment, and so differ in their respective practical judgements *without this difference, and so J's decision, being a matter of chance*. Both judgements may be justified and under the agents' rational control, even if they cannot be both true.

Let us discuss the 'Rolling-Back' version of the 'Mind' argument, in Van Inwagen's presentation thereof. On the basis of the undetermined character of Alice's choice, and of the 'Rolling-Back' thought experiment, Van Inwagen thinks we may plausibly conclude that, in the situation Alice faces, the actual result, namely her telling the truth, has the same probability, around 0.5, as the alternative result, namely her lying. So the fact that she actually tells the truth is a matter of chance and, we may add, not an adequate ground of her moral responsibility for that act. But why should we assign a probability of around 0.5 to each alternative? In making this move, Van Inwagen is taking a very external perspective of Alice's situation and thinking of indeterminism as a bottom feature of reality that transmits itself upwards, encompassing Alice's choice and action. But this view of both the situation and the role of indeterminism therein is not forced upon us. In connection with moral responsibility issues, we should rather make the effort to see things from the internal point of view of Alice herself when she faces the choice. In Van Inwagen's example, we are told next to nothing about Alice's point of view. We know almost nothing about her evaluative beliefs, especially concerning telling the truth and lying as kinds of actions, but this issue is central in our proposal. We may then imagine various possibilities.

Suppose, first, that Alice adopts a purely neutral stance, so that she does not think that telling the truth is either better or worse than lying. She does not think there is any difference in moral value between these two options either. Suppose also that, in this particular situation, considerations of another sort do not break the balance in favour of one of the options. This brings the situation close to a Buridan case. Here it is justified to hold the view that the probability of Alice's lying is more or less the same as that of her telling the truth, around 0.5. But this does not show that Alice lacks rational control over her choice for, in Buridan cases, it is a rational procedure to leave the decision to chance by, say, flipping a coin. Any of the alternatives will then be justified. An additional question is whether she truly deserves praise or blame for what she chose and did, and the answer depends upon Alice's having or not having ultimate regulative control over her neutral (and plausibly false) evaluative belief.

However, Van Inwagen's remarks that the situation was 'difficult' and that Alice 'seriously considered telling the truth, [and] seriously considered lying' (Van Inwagen 2000: 14) suggest that this reading of the example is not the one he has in mind. Alice is plausibly taken to assign moral relevance to her choice. Suppose that she has developed a system of evaluative beliefs that includes the general view that telling the truth, unlike lying, is morally good, though this value may be outweighed by other moral values, or even by non-moral ones, so that, in some situations in which this is the case, lying might be – even morally – justified. Suppose that her deliberation about the particular situation leads her to the view that, though there are some reasons for lying in this case, so that this option might have some degree of rational and even moral justification, these reasons do not seem sufficient to override the general evaluative belief that, *ceteris paribus*, telling the truth is preferable to lying. The practical judgement she arrives at, then, is that, in this particular situation, telling the truth is better than lying, and she decides and acts accordingly.

Are Alice's choice and action a matter of chance? What leads to this conclusion in the 'Rolling-Back' argument is that, given that the choice is undetermined, in some subsequent replays Alice will choose to lie instead, so that after a high number of such replays the ratio of each possible outcome with respect to the total number of replays would be seen to converge towards a certain value, around 0.5. This means that, in the actual situation, there was a probability of about 0.5 that Alice had lied, so that her telling the truth is a matter of chance. However, after trying to see things from Alice's perspective, and to bring her evaluative views and practical judgement into the picture, various steps in the argument, as well as its conclusion, seem much less forceful. It is now important to ask at what phase in Alice's practical deliberation the replays are supposed to start in the thought experiment. Van Inwagen supposes they start 'one minute before Alice told the truth' (Van Inwagen 2000: 14.) This is not much help, but it suggests that the replays are supposed to start at a time quite close to Alice's choice. Suppose they start at some moment between Alice's practical judgement and her decision. Then it is completely implausible to think that, in about fifty per cent of the replays, she will decide to lie and do so, unless we make the wild assumption that, in about half of the replays, she suffers an attack of weakness of the will. Suppose then, that the replays start just before Alice's practical judgement. Then it is also implausible to think that, with the same reasoning that led her to her actual practical judgement in the real situation, she would arrive at a contrary practical judgement in about half of the replays. To get the argument off the ground, we should suppose that the replays start before she begins to reason about the alternatives, or perhaps in the middle of that reasoning. Assuming some degree of bottom-up, quantum indeterminism in Alice’s brain, it may happen that in some of the replays some motives, related for instance to self-interest, appear to her as stronger than in the actual situation, thus leading her, in *some* of these cases, to a different process of reasoning and to a different practical judgement. It may also happen that in some replays she reasons more, or less, carefully, or takes other considerations into account, though this in itself does not mean that she will necessarily arrive at a different practical judgement: our justified practical conclusions enjoy some degree of protection against such changes. So, even assuming that these variations take place in some of the replays, it is still implausible, given Alice’s system of evaluative beliefs and her general mental traits, to think that the ratio of the two possible outcomes would take a value which suggested the conclusion that Alice’s actual decision is a matter of chance. Alice’s making the same choice as in the actual situation in subsequent replays is, we think, much more probable than her choosing otherwise.

What makes the sort of arguments we are considering seem unconvincing is the fact of facing them from the perspective of a reason-sensitive, top-down indeterminism, combined with a cognitive rather than volitive approach to moral responsibility. But indeterminism is still there. When Alice faces the choice between telling the truth and lying, what she will choose is undetermined. Her evaluative view and the data she has about the situation she faces do not determine her choice. She still has to reason about how the view applies to the particular situation, and about whether the latter falls within the scope of the former. It is in the nature of normative systems and practices that questions of this sort have open answers and leave room for different, and justified, judgements. Taking part in those systems and practices introduces into our life the reason-receptive indeterminism that is required for moral responsibility.

We may finally ask whether our evaluative beliefs, which, we have argued, make ultimate control over our choices and actions possible, are simply a matter of chance. It would be foolish to deny that the system of evaluative beliefs we end up having largely depends on luck, an expression that is shorthand for factors beyond our possible control. But this does not mean that our praise- or blameworthiness for our evaluative views is equally a matter of luck. Blameworthiness for our evaluative views may diminish or even disappear if we do not possess the required control over the factors that explain such views. To take an earlier example, this is what happens in the case of the landowner in Ancient Greece concerning his belief that slavery is morally permissible. Conversely, an agent’s praiseworthiness for her correct views may be increased by virtue of the unfavourable circumstances in which she formed them. In any case, given minimally favourable circumstances, concerning both our innate capacities and our environment, we can control our evaluative views and get justification for them by applying the relevant normative standards. As happens with our free choices and actions, the evaluative system we hold at a certain time is undetermined, but again this indeterminacy does not threaten our rational control, for it relates to the reason-receptive indeterminacy which, we have argued, is constitutive of normative systems and practices.

FINAL REMARKS AND CONCLUSION

We started this chapter with an attempt to diagnose the roots of scepticism about moral responsibility. A revision of the dialectic situation that results from the preceding four chapters showed that, while SMR's premise B (the incompatibility of determinism and moral responsibility) is supported by two main lines of argument, only one leads to SMR's premise C (the incompatibility of indeterminism and moral responsibility.) This line leads to premise C through the necessity of ultimate control for moral responsibility and the incompatibility of indeterminism and ultimate control. Now, since ultimate control seems clearly incompatible with determinism as well, then, if it is necessary for moral responsibility, the latter cannot exist. According to some authors, the incompatibility of ultimate control with both determinism and indeterminism is just a consequence of the fact that this condition is incoherent in the first place. This would seem to counsel a rejection of ultimate control as a condition of moral responsibility, as compatibilists have done. But we have taken a different route. Against compatibilists, we have defended the necessity of ultimate control for moral responsibility, understood as true desert; and, against the incoherence contention, we have tried to show that ultimate control is indeed possible. Now, since determinism seems clearly incompatible with that condition, given that there are no ultimate causal sources in a deterministic world, we have contended that ultimate control is nonetheless compatible with indeterminism. The main objection against this contention is that indeterminism turns our choices and actions into chancy and arbitrary occurrences, and so erodes our control (and a fortiori our ultimate control) over them.

We have tried to resist both lines of attack on ultimate control. We have argued for the hypothesis that what makes ultimate control falsely appear as an incoherent, and therefore impossible, demand is an approach to moral responsibility and its freedom-relevant conditions with an almost exclusive emphasis on choices and acts of will. And we have contended that it is also this approach that makes undetermined choices and actions falsely appear as matters of chance. On this basis, we have developed the makings of an alternative, libertarian approach, which instead underlines the central significance of cognitive factors and the existence of a form of control and true desert that does not give the will centre stage. We have especially insisted on the importance of evaluative beliefs for the possibility of moral responsibility. We have argued that, from this cognitive perspective, indeterminism can be seen to be receptive to rational control, against arguments to the contrary. This alternative approach, we have contended, clears the ground for an anti-sceptical stance towards moral responsibility, in that it undermines the central line of argument for SMR's premise C. However, humility is also an important virtue in philosophy, and it is especially important with regard to such recalcitrant questions as this book has been dealing with. More work should be done in order to fill in the details and reinforce the proposal against possible, and probable, objections. In the end, this proposal might well be shown not to work. In the meantime, however, let us enjoy the hope that it might succeed.

The cognitive approach to moral responsibility we have proposed in this chapter has important advantages over will-centred libertarian approaches, especially for what concerns rational control. However, we can also integrate important insights of both incompatibilist and compatibilist theories into this new context. We can accept, for example, with Kane, Benson and Richardson, that the freedom relevant to moral responsibility is not an inborn and indelible quality of human beings, but rather a contingent achievement that has to be acquired and perfected, and that can also be lost. We agree with Kane when he holds that moral responsibility not only concerns particular actions, but also the building of one's own character, of 'virtuous and vicious dispositions through one's own efforts, choices, and actions' (Kane 1996: 181), though we would also insist on the importance of cognitive virtues and attitudes, such as humility and respect for evaluative facts. And we can accept the relevance of external social and political factors to the development of one's freedom. We can allow for the possibility of a plurality of justified systems of evaluative views, as well as for its importance in the formation of a correct system of our own. Moreover, we can avail ourselves of compatibilist insights about the factors that increase, diminish or even preclude moral responsibility, against a rather inhuman view of human beings as unrestrictedly free and responsible agents and so unconditionally liable to punishment. And, of course, we can gladly greet the compatibilist insistence on rational conditions of moral responsibility.

NOTES

[1] The assumption that all control and responsibility requires the will's intervention is very widespread. For an example, consider the following text by R. Jay Wallace: 'Both negligence and recklessness can be taken to reflect *qualities of will*, as expressed in action, *and so* to be appropriate grounds for blame … Recklessness … involves a cavalier attitude towards risk that shows itself in the relation between one's choice and

one's awareness of the risk in acting on that choice … and so recklessness can itself be *a blameworthy quality of the will*. Negligence and forgetfulness are slightly harder cases, perhaps, because there may not even be awareness of the risks involved at the time when one acts negligently or forgetfully. Here one may have to trace the moral fault *to an earlier episode of choice* … In this way, negligence and forgetfulness may also be traced to a *blameworthy quality of will*' (Wallace 1994: 138–9, my emphasis). Examples of this assumption could be multiplied.

[2] In a recent paper (Zhu 2004) I have found some interesting remarks, which point in the direction of my defence of a form of control not based on the will. The seemingly paradoxical notion of 'passive action', which Zhu traces back to Frankfurt, involves the idea of control, though not voluntary. Non-interfering with a process can be a form of control over that process, whether or not it can rightly be taken to be an action. While driving, for example, there are several moments at which we do not *do* anything. Nonetheless, it seems right to say that at those moments we are still driving and in control of our car.

[3] I have argued elsewhere (Moya 1998) that justification implies (non-deterministic) causation. So, if a reason is to justify a decision or action, it has to cause it. On my view, in order to analyse reasons explanation correctly, we need not and should not conceive of the causation relation between reasons and action as independent of the internal coherence relationships between the (content of the) reasons and the (description of the) action that are essentially involved in justification. Justification already includes a causal requirement. This view, I would think, is in a better position

than Davidson’s to avoid problems such as epiphenomenalism of mental properties and deviant causal chains.