Did My Neurons Make Me Do It?

Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will

Nancey Murphy and Warren S. Brown

OXFORD
UNIVERSITY PRESS

Contents

This page intentionally left blank

Detailed Contents

7

Neurobiological Reductionism and Free Will

1 Prospect

There are two broad issues in considering free will, which we shall distinguish "prepositionally": there is freedom *from* (a wide variety of real or imagined) constraints, but also freedom *for* or *to*. The latter is variously described: to pursue the good, to act for reasons, for development of one's character, to pursue increasingly sophisticated understanding of the good.[1] We would add to this the capacity to *inhibit* one's propensities for action. These capacities closely match the cognitive capacities we listed in the previous chapter as prerequisites for morally responsible action. This should not be surprising, given the close connections between the problems of free will and moral responsibility.

Current literature on the free-will problem focuses largely on freedom *from*; that is, the major worry is (some form of) determinism. In this chapter we hope to do two things. One is to show that the preceding chapters have provided resources for dispelling the worry about one form of determinism. Then we raise questions about other aspects of the free-will problem as it is ordinarily formulated.

In section 2 of this chapter we shall report briefly on the status of the current philosophical debate regarding free will, and note other authors' judgments to the effect that it has reached a stalemate. We diagnose the stalemate as arising in large part from a narrow focus on the issue of determinism versus indeterminism.

[1] Cf. Alasdair MacIntyre, *After Virtue: A Study in Moral Theory*, 2nd edn. (Notre Dame, Ind.: University of Notre Dame Press, 1984), 219: "The virtues therefore are to be understood as those dispositions which will … furnish us with increasing self-knowledge and increasing knowledge of the good."

In section 3 we argue that the most pressing issue, instead, is reductionism: do agents in fact exercise downward control over their constituent parts (i.e., is reductionism false?)? Determinism/indeterminism at the neurobiological level is irrelevant, on our account, because downward causation amounts to selection among or constraint of lower-level causal processes, regardless of whether those lower-level possibilities are generated deterministically or randomly. In the next section we examine one of the most prominent accounts of free will, that of Robert Kane the "libertarian", and note the extent to which his position conforms to reductionist assumptions.[2] In section 5 we employ arguments regarding the nonlinearity of causation in complex systems and the teleological elements of human and animal behavior to question certain formulations of the free-will problem. In section 6 we examine Daniel Dennett's "compatibilist" position and conclude that, due to his reductionist approach, he is only able to argue for pseudo-responsibility and pseudo-free will.

In section 7 we briefly revisit the determinist issue, claiming to have defused the issue of neurobiological determinism and to have called into question the cogency of more general determinist theses.

In section 8 we raise the question of whether the moral reasoner we described in Chapter 6 (and for whose appearance the previous chapters laid groundwork) in fact meets the most important (and realistic) criteria for a person's possessing free will. We shall suggest that free will be seen as a holistic capacity of mature, self-reflective human organisms acting within a suitable social context.

2 The Stalled Debate

There is no such thing as *the* free-will problem. Over the centuries, philosophers and theologians have debated a number of problems that share a family resemblance. Ancient Greek dramatists explored the role of fate. In the early Christian era two problems arose. First, if God had predestined some humans to be saved (and by implication the others to be lost), as Augustine taught, is this reconcilable with anyone's freely choosing to

[2] This is an account of Kane's position in his very influential book *The Significance of Free Will* (Oxford: Oxford University Press, 1996). We are pleased that in his later writings he has endorsed conclusions very much closer to our own.

obey the will of God? The second problem (still hotly debated) is whether human freedom is reconcilable with divine foreknowledge. What these three have in common is that they are in one way or another opposing *some* concept of human freedom to *some* concept of determinism.

The major difficulty in addressing the contemporary free-will problem is the fact that the very meaning of "free will" is contested. In fact, even the preceding statement is contested: philosophers of the ordinary-language variety claim that we know quite well enough what we mean when we say that someone did something freely. It means that one was able to act as one chose, and was not, for instance, compelled to do it by having a gun to one's head.

For other philosophers, the question is not merely whether one is able to act as one chooses, but whether one is able to *choose* freely. But if all events are determined by prior causes, then must not human choices themselves be determined by prior causes? Thus, current philosophical literature is structured by the compatibilist–incompatibilist distinction: free will either is or is not compatible with determinism. The concept of determinism at play in current debates is often a general thesis such as the claim that every event or state of affairs must have a cause. There are two questions, then: Is the causal determinist thesis true? And if so, Is free will possible?

2.1 Interminable Arguments

Galen Strawson, in his article on free will in the new *Routledge Encyclopedia of Philosophy*, sees little chance of progress in settling this issue: "The principal positions in the traditional metaphysical debate are clear. No radically new option is likely to emerge after millennia of debate."[3] Similarly, Louis Pojman concludes a particularly lucid overview of the problem of determinism and free will with a confession of ignorance: "I do not know the answer to this enigma … [a] paradox which has, since the dawn of reflective thought, perplexed the very best minds."[4]

Strawson describes a cycle in free will/determinist debates. The motor that drives the cycle is the issue of moral responsibility: there are powerful arguments that seem to show that we cannot be morally responsible in

[3] Galen Strawson, "Free Will", in Edward Craig (ed.), *Routledge Encyclopedia of Philosophy* (London and New York: Routledge, 1998), iii. 743–53, at p. 749.

[4] Louis Pojman, "Freedom and Determinism: A Contemporary Discussion", *Zygon*, 22/4 (1987), 397–417, at p. 416.

the ultimate way we suppose, and these clash with powerful psychological reasons to believe that we *are* morally responsible. His account begins with the position most often in our day called *compatibilism*. Free will is taken to be compatible with determinism because in ordinary parlance, "free will" means being free from various forms of coercion or restraint.

However, the determinist objects that if all events are determined, then the distinction between actions for which one is and is not responsible (as in legal settings) must be invalid. An incompatibilist notion of free will is essential to make sense of the idea that we are genuinely morally responsible.

This step in the argument triggers what Strawson calls the pessimists' objection, the argument that *lack* of determinism cannot contribute to moral responsibility either. The pessimists therefore conclude that free will is not possible in the strong sense required for moral responsibility. And if this ultimate sort of responsibility is not possible, then punishment or reward can never be just.

Now, says Strawson:

> the argument may cycle back to compatibilism. Pointing out that "ultimate" moral responsibility is obviously impossible, compatibilists may claim that we should rest content with the compatibilist account of things—since it is the best we can do. But this claim reactivates the incompatibilist objection, and the cycle continues.[5]

2.2 A Clutter of Terms

In speaking of terminology it is important to note first that while "the free-will problem" and "the problem of the freedom of the will" are used as conventional designations for the issues at hand, they are no longer descriptively appropriate. The terms arose in an earlier era in which the soul was taken to have a variety of faculties. Plato distinguished three: reason, spirit (*thumos*), and appetite. Some scholars credit Augustine with the invention of our traditional Western conception of the will.[6] It was necessary in his theology to explain how it could be the case that a person knows the right thing to do, yet not do it. Given the shift to physicalism, it would be better not to speak of "the will" at all and to describe the problem as the question of free agency or free choice. Nonetheless we

[5] Strawson, "Free Will", 749.

[6] Cf. Phillip Cary, *Augustine's Invention of the Inner Self* (Oxford: Oxford University Press, 2000).

shall use the standard terminology. In light of our emphasis on avoiding the Augustinian and Cartesian picture of interior actors, we would want to insist that the agent or chooser is the whole person, neither a self or soul nor some bit of neural tissue inside the brain.

As noted above, the current discussion focuses on the libertarian and compatibilist positions. However, these are not exact contraries, as can be indicated initially by the following chart. That is, compatibilists believe that determinism is true and that we have free will nonetheless. Hard determinists believe that determinism is true, that free will is incompatible with determinism, and so we do not have free will. Libertarians believe that free will is incompatible with determinism but reject (complete) determinism.

	Determinism	Indeterminism
Compatibilism	Compatibilists	- - - -
Incompatibilism	Hard determinists	Libertarians

The picture is complicated, though, by Strawson's pessimists, who believe that free will is also *in*compatible with *in*determinism. We therefore need two categories of incompatibilist theses: one regarding determinism and one regarding indeterminism. Let us designate these as incompatibilism-d and incompatibilism-i. So the chart of possible positions actually looks more like this.

	Determinism	Indeterminism
Compatibilism	Compatibilists	- - - -
Incompatibilism-d	Hard determinists	Libertarians
Incompatibilism-i	- - - -	Pessimists

There are a variety of related terms in the literature. The incompatibilist–libertarian's criterion for choosing freely is being able to have chosen differently (and thus acting differently) than one in fact did. Thus, another term for libertarian free will is ''counterfactual free will''. An older term in the literature is ''liberty of indifference'', meaning that one has motivational equilibrium. Older terms for compatibilist free will are ''liberty of spontaneity'' and ''soft determinism'', the latter being contrasted with ''hard determinism''.

We suspect that the reason for the stalemate in free-will debates is the focus on the issue of determinism versus indeterminism. In the next section we consider one form of determinism, neurobiological determinism, and

argue that setting up the problem in this way is misguided, in that the real issue here is causal reductionism. In section 7 we remind the reader of our worries about the cogency and/or truth of more general determinist theses.

3 Defining the Determinist Threat

One problem with the compatibilist–libertarian debate is that the concept of determinism itself is too vague to be useful. One has to be more precise and ask what it is that is taken to determine human choices. We have set aside the ancient concept of fate, as well as the theological issues. In B. F. Skinner's day, social determinism was a pressing issue.[7] We have argued that dynamical systems exhibit a degree of autonomy from their environments. Juarrero says: "Because all self-organizing systems select the stimuli to which they respond, behavior constrained top-down is to that extent increasingly autonomous of forceful impacts from the environment. Self-organized systems act from their own point of view."[8]

In our own day the threat of biological determinism—genetic or neurobiological—has displaced environmental determinism. Genetic determinism can be set aside for two reasons. First, instances of human behavior that are in any way candidates for free choice (unlike height, for instance) are never perfectly correlated with genes. That is, identical twins show similarities in political attitudes, temperament, sexual orientation, and so forth, but the imperfect correlation means that genes are only part of the story. In fact, same genes plus same environment (i.e., for identical twins raised together) do not produce perfect correlations. A second line of reasoning is based on the total quantity of information contained in the genome versus the total amount of information that would be needed to determine the synaptic connections in any particular individual's brain. The genes fall short of this capacity by a factor of ten billion.[9]

[7] However, reductionists in sociology and psychology regularly translated environmental determinism into neurobiological determinism by arguing that the environment cannot affect the person unless the person is aware of it, and so it is the neural representation of the environment that is actually the causal player in the person's behavior—bottom-up, after all.

[8] Alicia Juarrero, *Dynamics in Action: Intentional Behavior as a Complex System* (Cambridge, Mass.: MIT Press, 1999), 249.

[9] There are known to be about 30,000 genes in the human genome. The usual estimate for the number of neurons in the human brain is 100 billion. Although the number of synapses per neuron

3.1 Defusing the Threat of Neurobiological Determinism

In contrast to genetic determinism, we have taken the threat of neurobiological determinism to be real and highly significant. However, this more specific focus actually misses the point in a subtle way. The issue is not whether the laws of neurobiology are themselves deterministic, but whether neurobiological *reductionism* is true. If we have made our case for the intelligibility of downward causation and its prevalence in shaping and reshaping the neural system, then the question of whether or not the laws of neurobiology are deterministic is actually not particularly relevant to the issue of free will. On our account of downward causation via selection or constraint, it makes no difference whether the laws of the bottom level are deterministic or not; higher-level selective processes can operate equally well on a range of possibilities that are produced (at the lower level) by either random *or deterministic* processes.

In contrast to the atomist–reductionist assumption of the modern world view, we have sketched a view of reality that recognizes the emergence, over the course of cosmic history, of new and more complex entities, *many of which possess new causal capacities*. (This is the rejection of atomist reductionism; see Ch. 2, secs. 1.2 and 1.3.) Some of these new capacities produce regular results, and thus their causal roles in the world can be modeled by deterministic laws. For example, the Hodgkin–Huxley laws that describe the transmission of nerve impulses are strict (deterministic) laws. If we give up the presupposition that the behavior of all higher-level entities must be deterministic, and simply look, we see that there are also complex systems (e.g., organisms) that depend on (largely determinate) lower-level systems but do not behave in regular (deterministic) ways—as we shall elaborate in the following subsection.

The widely accepted conclusion that the lowest level of the hierarchy of complex systems (the quantum level) is genuinely indeterministic should long ago have laid to rest the assumption of the determinism of higher levels based on the notion of (deterministic) bottom-up causation. There are a few cases in which the *indeterminism* of the quantum level works its way up to the macroscopic level. One example is when a single quantum

varies throughout the brain, a reasonable but rather conservative number per neuron would be 1,000. See E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 3rd edn. (Norwalk, Conn.: Appleton & Lange, 1991), 121.

event is sufficient to account for a genetic mutation.[10] But in most cases, the indeterminacy at the bottom gives rise instead to deterministic systems at the next higher level. If we have to look to see whether and where indeterminacy works its way up the hierarchy, should we not also have to look to see whether and where determinism works its way up to higher levels?

As we pointed out above (Ch. 2, sec. 1.3.2), it may in fact be the case that the first and most basic example of downward causation is the phenomenon of decoherence in quantum physics; that is, the effect of interaction or measurement on an otherwise indeterminate state. So the basic "causal structure" of the universe, all the way from the bottom to the top may be a dynamic interplay of downward causation from large webs of structures that have evolved over time with bottom-up constraints provided by the original lower-level constituents. Science has attempted and often succeeded in capturing the laws governing lower levels by sealing off simple systems from interactions with their environments. But this may in fact have resulted in a false conception of the very nature of reality.

3.2 The Irrelevance of Indeterminism in Animal Behavior

The way biology works on a number of levels is by means of luxuriant overproduction of a range of possibilities followed by selection among the variants on the basis of criteria pertinent to a higher level of organization. This was, of course, first recognized as the mechanism of the evolutionary process. Mutations (at the genetic level) and cross-breeding (at the organismic level) produce a varied population of offspring. Factors at the ecological level do the selecting. Another example is the immune system, which works by producing a vast number of antibodies, but only the ones that work are produced in great numbers. Neural connections are

[10] See Robert J. Russell, "Special Providence and Genetic Mutation: A New Defense of Theistic Evolution", in *idem*, William R. Stoeger, and Francisco J. Ayala (eds.), *Evolutionary and Molecular Biology: Scientific Perspectives on Divine Action* (Vatican City State and Berkeley: Vatican Observatory and Center for Theology and the Natural Sciences, 1998), 191–224. Other examples are superconductivity and superfluidity, which are essentially quantum effects at the macro-level, and various devices (such as the human eye) that register a quantum-level event (e.g., the reception of a single photon) and thereby amplify its effects. See George F. R. Ellis, "Quantum Theory and the Macroscopic World," in Robert J. Russell *et al.* (eds.), *Quantum Mechanics: Scientific Perspectives on Divine Action* (Vatican City State and Berkeley: Vatican Observatory and Center for Theology and the Natural Sciences, 2001), 259–91, at pp. 260–1.

overproduced in infancy and selectively maintained or strengthened only as they are used.

There are two parts to causal stories of this sort: (1) how the variants are produced (this was a problem plaguing evolutionary theory until the development of genetics), and (2) the basis upon which and the means by which the selection takes place. A fairly insignificant part of the story is whether processes that produce the variants are deterministic or indeterministic. For example, genetic mutation occurs in some cases due to factors at the quantum level (indeterministically) and in others due to macro-level (deterministic) processes.

Animal behavior needs to be understood on the same model. The picture we are prone to have in mind is an organism whose "default position" is inactivity. When it acts, the question arises as to what caused it to act and whether the action was "up to it" or "not up to it"—that is, was the behavior stimulated by something in the environment or by something internal to the organism. If the latter, then we want to know, could it have done otherwise?

A more accurate picture is of an organism that is constantly active (to some degree or other). Thus, the question is not what initiated any part of the behavior, but rather, what the factors are that modify ongoing behavior. As we pointed out (in Ch. 3, sec. 2.1.2), the action of even very simple organisms is the result of the emission of a variety of behavioral components that are either continued or altered on the basis of feedback from the environment and evaluation in relation to the organism's goals.

For example, the flying movements of a fruit fly can best be modeled by sampling from a probability matrix. At a particular moment in behavior, the ongoing context has activated a matrix of behavioral probabilities. The highest probability behavior is apt to be expressed, but a lower probability might also occur, particularly if the matrix of probabilities is rather flat. The matrix is influenced by current appetites, drives, avoidances, and previous learning, but also by feedback from immediate past behavior. The behavior that is expressed next is the outcome of sampling from this matrix. Once a behavior is sampled and tried out (e.g., turning 20 degrees left), the outcome is immediately evaluated, and the probability matrix modified (or not). Thus, behavior is continuously sampled, tried, evaluated, and modified, based on probabilities that are themselves constantly modified by experience and learning.

We suggested above (Ch. 2, sec. 5.4) that there are two possible meanings for "probability" is in this account. The major options would be to interpret them in terms either of relative frequency or of "propensities". The relative frequency option is inapplicable because an organism is never in exactly the same situation (internal states, environmental conditions, recent history of behavior) more than once. A propensity is defined as "an irregular or non-necessitating causal disposition of an object or system to produce some result or effect … usually conceived of as essentially probabilistic in nature".[11] Karl Popper regarded the existence of propensities as a metaphysical hypothesis needed to interpret probability claims about single cases.[12] If we give up on the notion that determinism and indeterminism are exhaustive categories, then perhaps strict *indeterminism* applies only at the quantum level (and in cases where macro-systems amplify quantum events); *determinism* applies largely to the realm of mechanical processes, and *propensity,* a genuinely distinct option, applies to much of organismic behavior. Juarrero interprets propensities in terms of dynamical attractors. She describes infants as coming into the world with a basic set of inherited propensities embodied in the dynamics of their biology, "the conditional probability distribution of the context-sensitive constraints that give the organism its identity".[13] But this landscape of attractors constantly changes in response to the infant's interaction with the environment.

The important point here, for present purposes, is that when top-down evaluation of emitted behaviors is the primary determinant of the larger course of the organism's activity, then it matters not whether the (bottom-up) production of behavioral variants is deterministic, indeterministic, or "probabilistic". Significant units of action are the "resultant" of three factors: (1) spontaneous emission of behavioral variants, (2) feedback, and (3) criteria for evaluation, both those built in by biology and those reshaped by action in the world. To put this in the language of structuring and triggering causes, the built-in goals are structuring causes; failure of an action to achieve a goal (a mismatch signal in MacKay's terms) triggers the emission of another behavioral possibility. In all but the most rudimentary organisms, the success or failure of an action changes the probability

[11] David Sapire, "Propensity", in Robert Audi (ed.), *The Cambridge Dictionary of Philosophy* (Cambridge: Cambridge University Press, 1995), 657.

[12] Karl R. Popper, *A World of Propensities* (Bristol: Thoemmes, 1990).

[13] Juarrero, *Dynamics in Action*, 176.

matrix according to which future behaviors will be emitted. This is a *downward* effect, from the system that is the whole organism acting in its environment, to the components of the nervous system that channel behavioral options.

If biological reductionism is false, and if most organisms are (to a degree) the causes of their own actions, and thereby top-down shapers of their own neural processes and structures, does this amount to free will? It would if "free will" meant nothing other than "not biologically determined". So we make the modest claim to have defused one possible element in the determinist's argument against free will. We turn now to ask whether our account of downward causation in organisms' behavior might shed light on any of the other debates in the current free-will literature.

4 Libertarian Reductionism

In this section we shall criticize the attempt to work indeterminism into an account of free will as being a direct result of the assumption of neurobiological reductionism. In section 4.1 we note the reductionist assumptions in Robert Kane's libertarian theory based on quantum indeterminacy. Our critique, however, will highlight valuable contributions, especially his claim that the essential meaning of free will is an agent's having *responsibility for his or her action*.

4.1 Robert Kane: Indeterminism in the Cartesian Theater

From Augustine through Descartes and beyond, the soul was understood to have a variety of faculties, one of which was the will. For an act to be free, it must be willed by the agent *and* the act of willing itself must not have a prior cause. Cartesian materialists substitute for a free *inward* act of will an undetermined event *in* the brain. Given the wide acceptance of indeterministic interpretations of quantum mechanics, an important move in recent free-will literature has been to attempt to use a quantum event to break the chain of causes in which the agent is assumed otherwise to be bound.

We are, in general, suspicious of such moves. First, it is a category mistake to identify a free choice with an indeterministic brain event. This is, in Donald MacKay's terms, an illegitimate mixture of the I story

with the brain story. It is the agent as a whole to whom the ascription of freedom applies.[14] Second, in line with our argument throughout this volume, we object to the reductionist assumption that the person's behavior must be the product of the working of his or her micro-level parts. Finally, as we argue in this chapter, the recognition of top-down causation makes lower-level determinism and indeterminism (largely) irrelevant.

Despite these objections, much can be learned, we believe, from examining the most sophisticated and well-respected attempt to secure free will via quantum indeterminacy. We shall argue that Kane had in fact already supplied most of the ingredients necessary for a viable account of free will in his theory published in 1996.

Kane places his account squarely in the libertarian camp. He defines free will as the power of agents to be the ultimate creators or originators of their own ends and purposes.[15] That is, if causal chains of actions are traced back to their sources, they must come to an end in the willings of the agents. If those willings are caused by something else—heredity, environment, God—then the ultimacy would not lie with the agent.

Kane takes it for granted that making room for free will, so understood, depends on there being occasions when the atoms "swerve", and this must take place in the brain.[16] Desires are among the inputs to practical reasoning; the outputs of will are choices, decision, intentions. There must be an indeterminate step somewhere between input and output, and this step Kane takes to be quantum-indeterminate events in the brain. The centerpiece of his work is to show how quantum events can be worked into an account of the freedom of the agent (Kane is very critical of the mistake of simply equating indeterminacy with free will).

Despite our diametric opposition to the centerpiece of Kane's work, we find ourselves in agreement with much of what he says. The most helpful elements of Kane's writings, if pulled into a different constellation, provide most of what is needed for an account of free will in terms of downward causation. The fact that Kane does *not* make this further step, this *Gestalt* switch, illustrates the extent to which reductionism retains its grip on the imaginations of even the most sophisticated thinkers in the field. Here are some of the major elements in Kane's writings that we endorse.

[14] See Ch. 1, sec. 3.1 above. [15] Kane, *Significance of Free Will*, 4. [16] Ibid. 17.

1. Kane, along with many others, recognizes that a crucial issue in the consideration of free will is that of character development. In many cases, what is wanted in the way of free will is an account of action flowing from the actor's character and deepest motives. But if one's character is entirely determined by something or someone other than the agent, then, Kane says, one may have some freedom of action but not freedom of the will.[17] So an important question is how one can become responsible for one's own character development. Kane speaks helpfully here of the necessity of there being, at points in the actor's development, "self-forming actions".

2. Kane, along with Harry Frankfurt, recognizes the importance of *levels* of desires, motives, goals. A particular choice *is* the agent's choice if the agent endorses it as compatible with her larger (higher) motivational system. This motivational system Kane calls the "self-network".[18] In the language developed in our Chapter 5, we would say that a particular neural event constitutes a choice (a mental event) only within the broader context of prior goals, deliberation, and remembered outcomes of action. In addition, following Kane, the choice is the agent's choice (rather than a mental event that simply happens to her) when it fits into the context of her self-network.

3. The argument between libertarians and compatibilists has largely focused on the question of whether the ability to have done otherwise is an essential condition for free will. Kane notes that a second criterion fueling incompatibilist intuitions is what he calls the ultimate responsibility criterion.

> The idea is this: to be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient reason (condition, cause, or motive) for the occurrence of the action. If, for example, a choice issues from, and can be sufficiently explained by, an agent's character and motives (together with background conditions), then to be *ultimately* responsible for the choice, the agent must be at least in part responsible, by virtue of choices or actions voluntarily performed in the past, for having the character and motives he or she now has.[19]

Kane argues that the ultimate responsibility criterion, rather than the alternative possibilities criterion, is the essential one. We agree that responsibility

[17] Robert Kane, "Some Neglected Pathways in the Free Will Labyrinth", in *idem* (ed.), *The Oxford Handbook of Free Will* (Oxford: Oxford University Press, 2002), 406–37, at p. 414.

[18] Ibid. 423–4. [19] Ibid. 407.

is conceptually the primary criterion for free will. We shall take issue only with Kane's use of "ultimate responsibility". We shall argue instead that an agent is never the sole cause of her own actions. Thus, the appropriate term is something like "primary responsibility". We come back to this in section 4.4.

Kane takes his account to be dependent upon further developments in brain science. The scenario that he hopes will be made more plausible by future research is this: when an agent faces a choice and there are conflicting motives on both sides, this puts the brain into a chaotic state—or in other terms, a state far from thermodynamic equilibrium—which in turn makes its processes (more) sensitive to quantum indeterminacies. Thus, the chaotic state serves as a possible route for the amplification of an indeterminate event at the quantum level.

The problem with such an account so far, as Galen Strawson has pointed out, is that an indeterminate event in the middle of a chain leading from desires to actions would seem to make it arbitrary rather than free. Kane's intention, then, is to show how an act whose causal pedigree contains an indeterminate event can still be (partially) a result of the actor's motives and character. Kane argues that while a quantum event may indeed be arbitrary, the choice that it triggers and the person's subsequent acting upon it are not arbitrary because it is in accord with *one of* the agent's prior motivational sets.[20] This, he argues, fulfills the widely accepted condition that an agent's reasons must play a role in the causal etiology of the action.[21]

4.2 Our Critique

The decision, as Kane describes it, is in accordance with (one of) the agent's motives, but, we argue, there needs to be more in order for the choice to be *hers*; it has to be *brought about by* her motives. Granted, the agent's motives in Kane's account play an *indirect* causal role: it is the conflict in motives that produces the chaotic state such that a macro-level transition can be caused by a quantum event.[22]

[20] Kane, *Significance of Free Will*, 130. [21] Ibid. 136.

[22] Motivation is an abstract concept covering a variety of neurobiological functions, mostly related to the activity of limbic brain systems. "These concepts include homeostasis, setpoints and settling points, intervening variables, hydraulic drives, drive reduction, appetitive and consummatory behavior, opponent processes, hedonic reactions, incentive motivation, drive centers, dedicated drive neurons (and

We might say that there is in Kane's account a downward *interpretation* of the quantum event in terms of the higher-level system of drives and goals, but not downward *causation* of the patterns of micro-level events by those higher-level goals. A second problem is that the opportunity for self-forming actions on Kane's account appears only in cases where the agent is faced with conflicting motives. We address this in section 4.3 below.

Our point here is not to pursue a critique of Kane's work for its own sake, but rather to illustrate the grip of reductionist thinking even in what is perhaps the most brilliant and judicious theorizing to be found in the libertarian camp. However, there is an interesting progression away from reductionism in Kane's thought. In *The Significance of Free Will* (1996) he quotes almost in passing a line from Gordon Globus that describes top-down causation in the brain. This is in the context of spelling out the reasons for viewing the brain as involved in chaotic processes. He speaks of the influence of the whole net of neurons affecting each individual node and of each node in turn affecting the whole.[23] There seems to be no further development in this direction in "Some Neglected Pathways" (2002). We suspect that despite publication in the same year, Kane's "Free Will: New Directions" was actually written later. Here, in response to the work of Timothy O'Connor, he writes:

> Indeed, I also believe that emergence of a certain kind (now recognized in self-organizing systems) is necessary for free will, even of the causal indeterminist kind that I defend. Once the brain reaches a certain level of complexity, so that there can be conflicts in the will of the kind required for [self-forming actions], the larger motivational system of the brain stirs up chaos and indeterminacy in a part of itself which is the realization of a specific deliberation. In other words, the whole motivational system realized as a comprehensive "self-network" in the brain has the capacity to influence specific parts of itself (processes within it) in novel ways once a certain level of complexity of the whole is attained. This is a kind of emergence of new *capacities* and indeed even a kind of "downwards causation" (novel causal influences of an emergent whole on its parts) such as are

drive neuropeptides and receptors), neural hierarchies, and new concepts from affective neuroscience such as allostasis, cognitive incentives, and reward 'liking' versus 'wanting'." See K. C. Berridge, "Motivation Concepts in Behavioral Neuroscience", *Physiology & Behavior*, 81/2 (2004), 179–209.

[23] Kane, *Significance of Free Will*, 130; referring to Gordon Globus, "Kane on Incompatibilism: An Exercise in Neurophilosophy", unpublished paper, 1995.

now recognized in a number of scientific contexts involving self-organizing and ecological systems.[24]

Our suggestion, then, is that recognition of the limitations of neurobiological reductionism and replacement with a recognition of the role of downward causation in cognitive and brain processes would allow all of these other valuable pieces of Kane's work to fall into a set of new relations. Kane postulates that presentation of a choice between two courses of action, each one desirable in its own way, moves certain relevant brain states far from equilibrium and sets up possibilities for chaotic processes. Let us grant this for the sake of argument. However, we would describe the process as follows. Our emphasis has been on self-transcendence—that is, on making one's plans, goals, motives, drives the object of second-order reflection. Donald MacKay offers an intriguing analogy. He describes what happens when a video camera is focused on its own monitor screen. If "we swing the camera to look at its own screen we find it is generating a pseudo picture. And if we zoom in to try to get more detail the whole thing goes haywire."[25] This is, in fact, a chaotic state, and it helps in imagining how intensely focused self-evaluation might result in radical changes in brain activity.

Kane suggests that in this chaotic state the system as a whole is moved in one of two directions by a random event that "breaks the tie" by definitively activating the neural network realizing one or the other of the agent's motivational sets. We suggest instead that the focusing on the problem moves the network far from equilibrium. (This may or may not involve a chaotic system and may or may not involve a definitive role for single quantum events.) In our scenario, the choice is determined not by the bottom-up action of one quantum event, but rather by selection of one of the possible options by means of a higher-order supervisory system.[26]

Kane emphasizes that self-forming choices are "value experiments", whose justification lies in the future.[27] In parallel, we emphasize that

[24] Robert Kane, "Free Will: New Directions for an Ancient Problem", in *idem* (ed.), *Free Will* (Oxford: Blackwell, 2002), 222–46, at pp. 241–2.

[25] Donald M. MacKay, *Behind the Eye*, The Gifford Lectures, ed. Valerie MacKay (Oxford: Blackwell, 1991), 10.

[26] Note that this is no homunculus in the brain. Rather, it is the whole person, in her environment, taking into account a broader range of factors than are involved in the lower-order system.

[27] Kane, "Some Neglected Pathways", 425.

such choices, for agents with sophisticated enough cognitive processes, result from running behavioral scenarios off-line and evaluating likely consequences in terms of the agent's previous experiences, goals, values, and moral conceptions. This is followed by enactment in the field of operation, which will supply ongoing feedback, changing synaptic weighting so as to enhance or inhibit the likelihood of the agent acting similarly in the future. In cases involving significant choices, this is the neural-cognitive-behavioral process of character formation.

Our picture is essentially different from Kane's in this regard: For Kane, an event at the bottom of the hierarchy of complex systems gets amplified by brain processes and moves the agent to follow one of the courses of action. Because it is in the context of the agent's prior deliberation, it is interpreted by the agent as *her* choosing that course of action. So long as it is consistent with or a reasonable extension of her self-network, "she will endorse that as *her* resolution of the conflict in her will, voluntarily and intentionally, not by accident or mistake."[28]

We object that it is still, on this account, the quantum event that is the primary cause of the agent's action, even though it may be endorsed as consistent with motives instantiated by higher-level (broader) neural networks. On our account, one of a variety of (lower-level) neural pathways for future action is selected on the basis of a higher-level supervisory system whose comparator recognizes a match with its goals and labels it "good to enact" rather than "bad to enact". So the difference is Kane's bottom-up causation, bit by bit, by means of random events having an effect over time on the agent's system of goals, versus our top-down selection of lower-level processes on the basis of goals shaped by previous actions in the world—especially by action within the social world, with the external scaffolding it provides in the way of symbolic representation, abstract concepts, and, especially, moral concepts.

4.3 The Ubiquity of Self-Forming Actions

We mentioned above that a defect in Kane's account is the restriction of self-forming actions to cases in which one faces a moral dilemma. We believe this is wrong, for two reasons. First, all complex organisms are constantly re-forming themselves. Almost every action has at least

[28] Ibid. 421.

minor effects on the organism's neural structures. This is the basis for the formation of habits—the "mental butlers" that relieve us of the requirement for conscious decisions about routine activities. Action results in feedback from the environment that promotes the development of somatic markers, which subtly guide future behavior. In a sense, these are all "self-forming actions", but we shall follow Kane (and Dennett—sec. 6) in retaining this term for actions that matter *morally*.

Kane's account of the narrow range of self-forming actions coheres with the narrow scope of much of modern ethical theory. James McClendon notes that many twentieth-century ethicists have restricted the scope of morality to decision making. The perfect expression is found in the claim of existentialist Jean-Paul Sartre that "true morality consists not in deciding this or that, but purely and merely in deciding. Decide, and you have acted morally, while not to decide is—bad faith."[29] McClendon traces this "decisionism" to a pervasive interiorization in Western self-understanding, particularly in Christian spirituality, that goes back to Augustine. Originally "decide" and its Latin cognate meant "to cut off"—to end a battle or a lawsuit by a decisive victory for one side. The emphasis on inwardness led to the application of the term to an interior "battle" in which the mind wavers until one "side" overcomes the other.[30] Unquestionably, McClendon says, "such inner struggle is a recurrent feature of human life. But in modern Christianity, the new exaltation of the will, together with the interiorization of the Christian life, made it seem that such struggles are not part but the whole of morality."[31] Despite its origin in Christian thought, decisionism was the core of modern ethics from the utilitarians and Kant to John Rawls.

More recently, under the influence of writers such as Alasdair MacIntyre, the focus of ethics has shifted to the development of character, understood as a continuous project situated in the context of social narrative.[32] We followed MacIntyre in locating the capacity for morality in exercising one's ability to evaluate that which moves one to act in light of a concept of the good. Thus, it is not so much deciding which way to resolve a dilemma that forms moral character as the second-order reflection on and evaluation of one's propensities, emotions, reasons, goals. This sort

[29] James Wm. McClendon, Jr., *Ethics: Systematic Theology, Volume 1,* 1st edn. (Nashville: Abingdon Press, 1986), 56–7.

[30] Ibid. 58. [31] Ibid. [32] See esp. MacIntyre, *After Virtue*.

of reflection does not depend on an interior "battle"; it is more likely to occur as a result of social interaction. The development of moral character involves the development of the *habit* of such self-reflection. Consequently, the scope of moral action includes not only actions undertaken on the basis of conscious decisions (let alone agonizing decisions), but also the much more frequent actions undertaken automatically in light of moral evaluations made in the past. What matters most, McClendon claims, is not deliberate decisions, if any, but "unreckoned generosity … uncalculating love … 'aimless' faithfulness".[33]

Our second critique of Kane's account of self-forming actions is related to the first, and applies equally to MacIntyre's definition of moral responsibility. Both authors seem to be assuming a picture of human nature in which inactivity is our "default" condition and action begins as the result of some inner event. MacIntyre speaks in terms of "that which moves one to act"; Kane focuses on the quantum event that becomes magnified into a decision. We have emphasized throughout that organisms' "default" condition is constant activity. So interesting questions about human actions will be questions not (usually) about what triggered a particular act but rather about the criteria for evaluating a variety of branching possibilities for an ongoing series of actions—in dynamical terms, the factors reshaping the ontogenic landscape representing the possibilities of future actions. This is another route that leads to the conclusion that character formation must be (primarily) a matter of the long-term accumulation of evaluations of one's own criteria for evaluation of action. If we are always already active, like other organisms, then the moral challenge, as both the parent and the spiritual director know, is to develop the habit of stopping long enough to reflect on what we are doing and why.

4.4 Ultimate versus Primary Responsibility

Despite the fact that Kane terms his primary criterion for free will "ultimate responsibility", he recognizes that self-forming actions never totally re-create the person's character. Thus, as is recognized in legal settings, one's responsibility for an act undertaken in line with one's character is always only partial:

[33] McClendon, *Ethics*, 59.

I think what motivates the need for incompatibilism is an interest in ... a control related to our being to some degree the ultimate creators or originators of our own purposes or ends and hence ultimate "arbiters" of our own wills.[34]

We agree that humans are never entirely responsible for their own characters. We come into the world with some degree of initial biological (genetic) predetermination. As with other organisms, we are always already active due to this innate biological machinery. We try out various actions and modify our behavioral tendencies based on feedback. The maturation process is that of slowly developing higher-order evaluative systems that nest and modulate the systems that control our biological processes, and having built into our nervous system maps of how the world works. This action–feedback process involves increasing susceptibility to social influences. However, the childhood task is not only social adaptation, but also the development of autonomy. In biological terms, this involves development of capacities for intentionally directed action. In social relations, an important step is the development of a theory of mind, which allows one to distinguish one's own perceptions and desires from those of others and to predict the likely mental lives of others.

As we emphasized in Chapter 6 and in the previous subsection, the most important step, which separates (older) children from animals, is development of the ability to evaluate one's own actions and, especially, one's *reasons* for action. This begins with evaluation in light of conformity to parents and peers, but ultimately in terms of reasons, goals, and values that can be expressed only via symbolic language.

Graphically, we might represent human development as a trajectory in a two-dimensional space. One axis is from total biological determinism to total biological flexibility; the other is from complete social determinism to total social autonomy. At birth we are neither determined by society nor autonomous; we have to develop the capacity for social control *before* we can begin to establish our autonomy. Also, given the fact that our bodies tend to fail in old age, there is often a return toward biological determinism as well as loss of autonomy. Thus, a typical trajectory might be represented by the solid line in Figure 7.1.

The dashed line represents what some philosophers would see as the ideal trajectory. However, we emphasize that this process does not *and should not*

[34] Kane, "Some Neglected Pathways", 432.

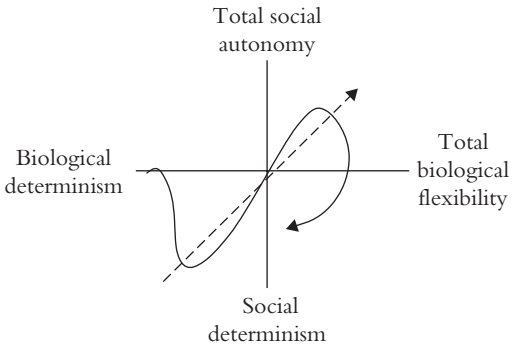

Figure 7.1. Trajectory of a human life (*solid line*) with respect to biological determinism–flexibility (horizontal axis) and social determinism–autonomy (vertical axis). The dashed line represents the idealized trajectory of some philosophers toward complete freedom from either biological or social determinism.

end with total flexibility and autonomy. Both biology and human culture have evolved to promote human flourishing. The ability to act contrary to both our biological nature and all social constraints can be predicted to lead (often enough) to disaster. Juarrero, noting that we are in part a product of "a complex dance" between our innate endowments and an already structured physical and social environment, says: "We are enmeshed in a fabric of time and space, which we unravel at our peril."[35]

For a somewhat whimsical depiction of the results of pursuing the goal of total autonomy, we return to our pacifist in Chapter 6. Let us return to the point at which he has been convinced by the teaching and example of the Jesuits to become a pacifist, but through an act of self-transcendence has come to suspect that his decision was determined by a need for social approval. The advocate of libertarian freedom might well raise the objection that surely something *caused* him to question the source of his pacifism, and if so, then again he is not free. To pursue this issue, *let us grant that there is a cause*, and let us further suppose that the causal factor is, in fact, his being given a book that convinces him of the ultimate importance of libertarian free will. Let us also suppose that this *causes* the emergence of a higher-level supervisory system that is *determined* to make freedom from biological and environmental determinism its highest priority. Detecting the influences of his community on the decision to become a pacifist, he immediately

[35] Juarrero, *Dynamics in Action*, 260.

abandons his pacifism. But now he recognizes that he was caused to make this change by the book he was given, and so his highest-level supervisory system requires rejecting his rejection of pacifism. But acting contrarily to the authority of the book is equally to be influenced by the book, only in a negative way. So what next?

Kane, of course, also rejects such a notion of free will. So where does our disagreement lie, since his ultimate-responsibility criterion in fact requires only that the person have some responsibility for his own character-forming decisions? We quoted him above saying that the person "must have been responsible for choices or actions *voluntarily* performed in the past".[36] But "voluntarily" here means without proper cause, and thus the central point where Kane believes himself compelled to turn to indeterminism—in a linear chain of determined events, there must be an occasional chink. This brings us to our next topic: a critique of the linear picture of causation implicit in this demand.

5 Questioning the Regress Argument

Peter van Inwagen is credited with one of the sharpest arguments for incompatibilism. Stated briefly:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[37]

5.1 The Nonlinearity of Human Responsibility

Note both the linear model of causation presupposed in van Inwagen's argument as well as the dichotomous option: up to us or not up to us. Our Chapter 2 was devoted to an exploration of the many reasons for rejecting a linear model, especially one in which it is assumed that each event in the linear sequence is simply determined by the previous event along with a relevant law of nature, as in Figure 7.2.

[36] Kane, "Some Neglected Pathways", 407; italics added.
[37] Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), 16.

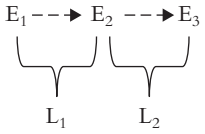

Figure 7.2. A linear model of causation in which each event (E_1, E_2 …) is determined by a prior event along with a relevant law of nature (L_1, L_2 …).

We first noted the role of boundary conditions in co-determining the effect of a given event. Pictorially this is easiest to represent by incorporating Fred Dretske's structuring causes into the picture. Many events occur because of the intersection of two streams of causal trajectories, as in Figure 7.3. The structuring causes that create conditions in organisms such that a given triggering cause is able to produce a given effect are in the first instance biological.

The next step in the defeat of simple linear models of causation is the recognition of the role of feedback. When an organism is triggered to act, its actions have consequences in the environment. Positive and negative consequences are fed back to the organism, often resulting in restructuring the organism itself. At this point we need to refer (one last time!) to Donald MacKay's diagram involving an organizing system, effectors, receptors, and a comparator (Fig. 2.3). Initially O is structured largely by genetics, with limited capacities to respond to triggering causes from the environment. Again, it does not matter whether the variety of responses is hard-wired or produced in some indeterministic or probabilistic manner. What is important is that positive or negative results in the field of operation result in the restructuring of O. At this level of complexity we have an organism restructured by experience, so it still seems appropriate to say that what it does is "*not* up to it", even if its responses were emitted by means of genuinely indeterministic processes—again we see the irrelevance of determinism versus indeterminism.

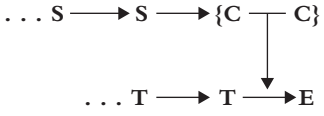

Figure 7.3. A series of structuring causes leading up to a condition such that a triggering cause is able to cause an effect.

We then considered systems capable of evaluation and modification of their own goals as a result of their actions. This is the beginning of a self-modifying system. From this point on, we have a system whose activity is in some minimal sense "up to it".

We added, then, the capacity for higher-order self-evaluation, the capacity to run behavioral scenarios off-line so that behavioral propensities could be restructured by imagined positive or negative feedback, and, finally and most importantly, the ability to evaluate behavior and one's own propensities in light of abstract symbols—to subject one's behavior to reason.

This process is vastly too complex to represent diagrammatically. The important point is that it is not a linear process, since the feedback from the environment *and from the person's own higher-order evaluative processes* is constantly restructuring the structuring causes themselves. We have to imagine countless iterations of such causal loops, the consequence of which is that the structural causes are to a greater and greater extent the consequences of prior actions and *evaluations* of the organism.

If we recognize (along with Kane) that the dichotomy "up to us, not up to us" is entirely unrealistic, then the important issue is that a (fuzzy but real) boundary is crossed at some point, so that some decisions and actions are to be attributed more to the agent than to the triggering cause together with structuring causes put in place by biology and the environment. In everyday life, in moral discourse, and in legal practices, we in fact judge whether people have acquired the capacity for responsibility, and by that we mean not total responsibility but rather that they bear the *primary* responsibility for their actions and decisions. Dynamic systems act from their own point of view, but based on history.

5.2 From Mechanism to Teleology

A second problem with the image of ourselves as having our decisions determined by a long series of causes reaching back to the beginning of the universe is failure to recognize that properly equipped organisms are teleologically directed entities. The modern scientific revolution began with the rejection of Aristotelian categories, among which were teleological explanations of natural phenomena. Thus, we have been set up by several centuries of history to view mechanistic causes as utterly different from teleological causes (if in fact there be any of the latter at all). We have

argued, in contrast, that proper design of mechanisms endows them with the capacity to act teleologically. We saw this even in simple devices such as thermostats. Their ability to pursue goals is not in contrast to their obedience to mechanical causes, but is enabled by it.

Thermostats are not particularly interesting here, because their goals are set by their designers; nor are the simplest organisms, whose goals are fixed by inheritance. The really interesting cases appear in conscious organisms, and especially in humans, whose imaginations and symbolic abilities allow them to imagine future scenarios, evaluate them, and shape their behavior accordingly. Language allows us to build up elaborate models of possible futures, which, along with finely modulated and flexible descriptions of motives, provide "final causes" for behavior.[38]

6 Daniel Dennett's Compatibilist Reductionism

Daniel Dennett's recent *Freedom Evolves* has drawn a great deal of attention, but reviews by philosophers have expressed deep disappointment with the book, dismissing it as just another version of compatibilism.[39] There are striking parallels between Dennett's book and our work in this volume (some not accidental, since we have made use of his earlier work in *Elbow Room*). In this section we shall first spell out the most significant parallels between our account and Dennett's, and then point out some of the ways in which Dennett encourages his critics' conclusions that what he offers in the way of free will is *not* worth wanting. We then attempt to sort out the underlying reasons for Dennett's book appearing to be so much like ours and yet, we shall claim, utterly different in essence. In the process, we hope further to muddy the libertarian–compatibilist waters.

6.1 Striking Parallels

There are structural parallels between Dennett's treatment and ours. Both begin with a story of increasing flexibility and use of information, leading

[38] Both Juarrero and Alwyn Scott favor reinstating all of Aristotle's four causes. Scott (in personal communication, 16 Aug. 2004) identifies triggering causes as efficient, structuring causes and boundary conditions as formal causes. Juarrero describes an attractor as a rudimentary precursor of a final cause (*Dynamics in Action*, 127).

[39] Daniel C. Dennett, *Freedom Evolves* (New York: Viking, 2003).

to degrees of self-direction in the animal world. Both emphasize the role of language in human affairs, accounting for the development of selfhood, self-transcendence, moral responsibility, and, finally, free will.

Dennett examines the development in living organisms of what he calls ''evitability''. Determinists are often, essentially, fatalists—concluding that whatever happens happens *inevitably*. Dennett examines ways in which organisms are equipped to evade outcomes that go against their interests by using information to guide the increasingly flexible behavioral repertoires one finds as one goes up the ladder of complexity. The result is organisms such as ourselves who are ''virtuoso avoiders, preventers, interferers, forestallers'' of happenings that do not serve their interests.[40]

Dennett emphasizes the critical role of language. It is language that makes hominids into persons (p. 173). Language, when installed in the human brain, creates the capacity for morality (p. 260). The shift from the amoral unfreedom of an infant to moral agency is gradual and involves environmental scaffolding and bootstrapping (p. 273). It is our practice of asking children about the reasons for their behavior that turns them into selves (p. 273). Engaging in this social practice allows us to develop the capacity for examining our own reasons (p. 249). We develop the capacity ''for monitoring not just the results of [our] action, but of [our] prior evaluations and formation of intentions as well'' (p. 248). While Dennett rejects Kane's turn to indeterminism to secure free will, he does make use of Kane's concept of self-forming actions, and notes that such an evaluation of one's reasons is a prime example of a self-forming action (p. 251). We are more readily redesigned—by ourselves—than any other organism on the planet (p. 277).

In almost the same terms as MacIntyre, Dennett states that our capacity to engage in the practice of asking and giving reasons, when directed toward ourselves, ''creates the special category of voluntary actions that sets us apart'' (p. 251). Dennett rejects the notion that freedom is an all-or-nothing concept, and reminds us of the fact that we have usable, if not infallible, cultural resources for discriminating degrees of responsibility (p.291). One often *becomes* responsible by *taking* responsibility—this is another crucially important sort of self-forming action (p.302). This parallels our emphasis on using one's capacity for self-transcendence in order to free oneself from

[40] Daniel C. Dennett, *Freedom Evolves*, 54. Parenthetical page references that follow refer to this text.

undue social pressure. There is no need for self-forming actions to be undetermined (p.302); the incompatibilists' arguments commit a misstep by pushing the causal chain preceding a person's action too far back in time. ''Events in the *distant* past were indeed not 'up to me,' but my choice now … is up to me because its 'parents'—some events in the *recent* past, such as the choices I have recently made—were up to me (because *their* 'parents' were up to me), and so on, not to infinity, but far enough back to give my *self* enough spread in space and time so that there is a *me* for my descriptions to be up to!'' (pp. 135–6). Throughout the book Dennett argues against taking the self to be some small operator in the brain (e.g., p.302). He states that this *I*, the larger, temporally extended self can control what goes on inside (p. 253).

Dennett's reviewers accuse him of having broken no new ground. Two critical reviewers are Galen Strawson and Jerry Fodor. Fodor says that Dennett's compatibilism, like other versions, turns out to be ''a sort of Chinese lunch: there's the lurking sense that what you got isn't quite what you ordered, and half an hour later you're hungry again''.[41] Fodor and Strawson both conclude that Dennett should bite the bullet and acknowledge that ''radical freedom'' (Strawson),[42] ''metaphysical freedom'', ''freedom *tout court*'' (Fodor) is impossible.

To a great extent these reviewers' reactions are simply a result of their accepting the libertarian–compatibilist dichotomy. Yet Dennett gives his detractors plenty of ammunition to argue that he has engaged in an exercise of ''bait and switch''.[43] Regarding intentionality, Dennett's claim is not that humans (and some animals) actually have beliefs, desires, intentions, but rather that we are pragmatically justified in employing the ''intentional stance'' (i.e., attributing to them beliefs, etc.). Since we cannot understand human behavior by means of an account of the calculations made by the myriad tiny robots in the brain, and because the ascription of beliefs, desires, intentions *does* permit us to predict behavior, we are justified in assuming the intentional stance with regard to our fellow humans. Fodor labels this an instrumentalist view of agency: ''there's no more to a creature being an

[41] Jerry Fodor, ''Why Would Mother Nature Bother?'', *London Review of Books*, 6 Mar. 2003, 17–18, at p. 17.

[42] Galen Strawson, ''Evolution Explains It All for You'', *New York Times Book Review*, 2 Mar. 2003), 11.

[43] Fodor, ''Why Would Mother Nature Bother?'', 17.

agent than its behaving like an agent.''[44] ''I strongly suspect that, though it may be heuristically useful for predicting behavior, the kind of agency that instrumentalists can certify is useless for explaining it … heuristic fictions don't cause anything.''[45]

In Dennett's account of altruism he describes how ''benselfishness'' can emerge as a result of recognizing, first, the social pay-off of being perceived as a good person and, second, that the easiest way of being so perceived is actually to *be* a good person.[46] Thus, reviewer Adam Schulman rejects Dennett's claim to have explained morality: Dennett has given an account only of ''pseudo-altruism'', whereas morality requires doing what is right for its own sake, not merely for selfish reasons.[47]

6.2 The Deep Difference: Reductionism

Is it possible to make a case for moral responsibility and free will, or only for pseudo-responsibility and pseudo-freedom? We have been suggesting that the answer lies in the concept of downward causation. Dennett looks at all of the right ingredients for an account of free will, but his goal is to *interpret* them all as products of bottom-up causation. Dennett has rightly concluded that determinism versus indeterminism in science is not the critical issue (cf. p. 306), but he does not distinguish between the determinism/indeterminism issue and that of causal reductionism.

We claimed at the outset that an account of humans that does not reject Descartes's view of bodies as machines is bound to be reductionistic. Dennett explicitly describes a human as an organization of a ''trillion robot teams''; we are each ''*made of* mindless robots and nothing else'' (p.2). This mind-set is the root of the reviewers' (accurate) perceptions that Dennett can account only for the *appearance* of intentionality, altruism, responsibility, and freedom. To illustrate, consider Dennett's position on language. This is taken from his earlier *Elbow Room*, but relevant here because so much in his current account hinges on human language use.

Dennett argues that our brains only appear to be ''semantic engines'', that is, meaning manipulators. Rather, being physical machines, they can only be ''syntactic engines'', responding to the structural or formal properties

[44] Fodor, ''Why Would Mother Nature Bother?'', 17. [45] Ibid. 18.
[46] Dennett, *Freedom Evolves*, ch. 7.
[47] Adam Schulman, ''Why Be Good?'', *Books in Review*, May 2003, 71–3.

of language.[48] Because meaning does not reside in the physical features of stimuli, no physical process could distill or respond to it.[49] However, Dennett argues, human capacities for meta-level pattern recognition allow for so much fine-tuning of our linguistic behavior that eventually "the overpowering 'illusion' is created that the system is actually directly responding to meanings".[50]

We take it that Dennett is making a point comparable to our Wittgensteinian rejection of understanding meaning as dependent solely on an inner "aha!" But lacking an alternative account of meaning in terms of language games and "grammar"—in terms of action in the social world—he concludes that there must be no meaning at all.

Dennett's position on language is an appropriate *entrée* for "diagnosing" his entire project, since he describes it as the first stable conclusion he reached in his philosophical career and as the foundation of everything he has done since then.[51] It is the source of his views of the nature of consciousness: "The appreciation of meanings—their discrimination and delectation—is central to our vision of consciousness, but this conviction that *I*, on the inside, deal directly with meanings turns out to be something rather like a benign 'user illusion'."[52] The denial of our ability to *understand* the meanings of words is also the foundation of his claim regarding the intentional stance. When he says that humans have beliefs, he means only that their behavior can be predicted by treating them as though they do. In this sense, "even lowly thermostats have beliefs."[53] Furthermore, if people act only *as though* they have beliefs, then it would seem to follow that they act only *as though* for a reason—for example, *as though* for the good of another person (pseudo-altruism)—and only *as though* taking responsibility for their actions.

We detect a profound irony here. Dennett is the inventor of the term "Cartesian materialism", which he uses to pillory the cognitive scientists and neuroscientists who expect there to be a place in the brain "where it all comes together", as in Descartes's mental theater.

[48] See Juarrero, *Dynamics in Action*, ch. 11, for a solution to this problem.

[49] Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, Mass.: MIT Press, 1984), 28.

[50] Ibid. 30.

[51] Daniel C. Dennett, *Brainchildren: Essays on Designing Minds* (Cambridge, Mass.: MIT Press, 1998), 357.

[52] Ibid. [53] Ibid. 327.

His rejection of a semantic account of language is specifically the result of his rejection of a neural substitute for the homunculus, who is the one on the "inside" who "gets the meaning". We have used "Cartesian materialism" more broadly to designate views of the mental that fail to take account of context: action–feedback–evaluation–action loops in the social environment. Thus, our account of the meaning of language depends not only on dynamical semantic networks *in* the brain, but especially on the *use* of language in the social world.

Our account of meaning depends on top-down causation at two levels. First, to understand a word *symbolically* is to have a hierarchically ordered semantic network such that higher-level categories (e.g., *food*) constrain the use of lower-level terms ("banana"). Because there is no exhaustive definition of food (as for many concepts), the system cannot simply be constructed bottom-up. The studies we reported on language learning in chimpanzees showed that while the first set of food items was learned by rote, the chimpanzees had, in Wittgenstein's terms, to learn "to go on", to extend the categories to incorporate new items. There did appear to be a point when the animal "got it".[54] This involved (as Dennett would grant) a higher-order recognition of patterns (edibles versus drinkables), but the ability to extend the categories was also dependent on both embodiment and interaction with the world—the downward influence of participation in language games. The category of food is determined not only biologically but also culturally.

So Dennett is right that higher-order pattern recognition is essential to language learning. He is also right that if knowing the meaning of a word could only be the result of better and better recognition of patterns among formal features of words and sentences, then it could never happen; machines can manipulate symbols but cannot understand them.[55] But knowing the meaning of language is a social achievement; it is the ability to use language appropriately in social interactions. It is also a result of the patterns of resemblance and causal and other connections that go into the use of icons and indices, respectively, as Deacon argues (see our Ch. 4, secs. 2–3).

[54] In Juarrero's terms, this is a catastrophic reconfiguration of the semantic and syntactical dynamic landscape; *Dynamics in Action*, 173.

[55] See Lakoff and Johnson's critique of the formal-semantics paradigm in linguistics, quoted in our Ch. 4, sec. 4.1.

We noted above that Dennett takes his position on language to be the source of his positions both on the intentional stance and on consciousness. We have argued that Dennett's account of language does not amount to (genuine) symbolic language. But we do not take the absence of symbolic language (as in higher animals and young children) as a reason for denying that they have (genuine) beliefs and other intentional states. Following Fred Dretske, we treated beliefs (in Ch. 5) as structuring causes of behavior. In this sense, human and animal beliefs are on a continuum with other sorts of information built into an organism's neural system (e.g., by genetics or classical conditioning).

Despite the continuities, however, an extremely important threshold is crossed when consciousness appears (see Ch. 3, sec. 5). Differentiating our position from Dennett's is complicated throughout by his stance on consciousness. We confess that we are unable to decide whether he denies the existence of consciousness or merely argues against misleading accounts of the nature of consciousness. Galen Strawson claims that "Dennett continues to deny the existence of consciousness, and continues to deny that he is denying it".[56] It may be significant in this regard that Dennett ends his *Consciousness Explained* with a section headed "Consciousness Explained, or Explained Away?"[57] Unfortunately, this brief section did not remove the question mark for us.

Our conclusion, then, is that Dennett's world view is largely Cartesian: his account of bodies is mechanistic; he works with a strictly bottom-up account of brain functions; he has an inadequate appreciation of the fact that the mind is essentially social and engaged in action. As noted above, Dennett remarks that "*I*, the larger, temporally and spatially extended self, can control, to some degree, what goes on inside",[58] but this does not seem to reflect his settled views of how organisms work, and especially does not take into account the downward causation from the social environment. So despite the parallels noted above, Dennett's book is, in the most important sense, as different from ours as could be. Our goal throughout has been to show where and why reductionism fails. Dennett's is *specifically* to provide a reductionist account of human phenomena, including "free will"; but the result is a very imaginative account of how

[56] Strawson, "Evolution Explains It All for You", 11.
[57] Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown, and Co., 1991), 454–55.
[58] Dennett, *Freedom Evolves*, 253.

complex machines could *appear* to have language, beliefs, morality, and free will.

7 Determinism Revisited

One thesis of this chapter has been that contemporary free-will debates have stalled due to their focus on the issue of determinism. Determinism *would be* the crucial issue if the early modern atomist–reductionist picture were true. That is, if the causal capacities of complex entities were nothing but the combined causal effects of the entities' constituents, and if the most basic constituents operated according to deterministic laws, then it would indeed seem to be the case that humans could do nothing other than what their atoms, in aggregate, do. We would have what Stuart Goetz calls bottom-to-top determinism,[59] in contrast to our common-sense notion that macroscopic entities are causes in their own right.

We have argued that this picture is wrong on three counts. First, it is widely accepted that the "atoms" (in the philosophical sense) do not behave deterministically. Second, it is becoming more and more widely recognized that complex dynamical systems can exhibit new sorts of causal capacities not found at the level of their constituents. We have emphasized, among these, sentience, goal seeking, consciousness, acting for a reason, and self-evaluation. Third, we have argued that higher-level systems exert downward effects on their constituents via selection among possibilities generated randomly, probabilistically, or according to deterministic lower-level laws. Thus, we have judged the issue of bottom-to-top determinism to be (largely) irrelevant to the free will problem.

This, of course, is only one sort of determinist threat to free will; there is also the form of determinism that Goetz calls past-to-present determinism. However, we have already noted the difficulties in making sense of a general past-to-present determinist thesis (Ch. 2, sec. 1.3.4). If determinism is defined as the *general* thesis that every event has (or *must* have) a cause, then at this point in intellectual history it is a very debatable metaphysical claim, and there are a variety of reasons for thinking that the issue could never be settled in favor of determinism (apart, perhaps, from a theological perspective).

[59] Stuart Goetz, "Naturalism and Libertarian Agency", in William Lane Craig and J. P. Moreland (eds.), *Naturalism: A Critical Analysis* (London and New York: Routledge, 2000), 156–86, at pp. 167–8.

8 Constructing a Concept of Free Will

Our goal in this section is twofold. First, we have attempted to show that the critical issue in the debate is one's ability, as a whole person, to exercise downward control over one's own behavior, cognition, and neural processes, and in so doing to become a causal player in one's own right. But if animals do this as well as humans, and if animals are taken *not* to possess free will, then there is the further (conceptual) question of the conditions under which such downward control in fact constitutes free will. The longer history of free-will debates offers a variety of options for analyzing the concept. We shall examine several of these here and relate them to the cognitive capacities incorporated into our account of morally responsible action (in Ch. 6). The goal will be to show that most of these traditional requirements are satisfied by our morally responsible actor. A second purpose, though, is to provide something of a corrective to these traditional concepts themselves in light of (what we hope is) a neurobiologically realistic account of human cognition and behavior.

8.1 Alternative Conceptions

We shall consider here (1) free will conceived of as acting for a reason; (2) various versions of free will as autonomy, distinguished by what one is taken to require autonomy *from*; (3) Harry Frankfurt's ''hierarchical mesh theory''; and (4) free will as dependent on ''agent causation''. We shall then raise the question of whether the conditions for morally responsible action are in fact the same as those for free will.

8.1.1. Freedom as Acting for a Reason Immanuel Kant is largely responsible for the understanding of free will as acting for a reason. Kant identified the moral law with what reason demands, and further identified freedom with the conformity of the will to the moral law. Kant's sharp distinction between acting for a reason and being caused to act was based on his distinction between noumena (things in themselves) and phenomena (things as they appear to us). The concept of causation applies only in the phenomenal world. The self is noumenal, and thus immune from causal forces.

Even though Kant's noumenal–phenomenal distinction is no longer influential, there are still many thinkers who see reason and causation as irreconcilable. The point of Chapter 5 was to call this assumption into

question, and to begin to spell out in neurobiological terms how "reason gets its grip on the brain"—that is, how reasons become causal players in human affairs.

Alasdair MacIntyre has argued that the absence of language is no reason to withhold the attribution of reasons to animals.[60] Nonetheless, there is a marked increase in freedom from biological and environmental demands that comes with symbolic language, and particularly with abstract concepts such as moral concepts.

It has been common throughout the Western tradition to contrast acting for a reason with acting under the influence of passion. This goes back at least to Plato, who postulated three aspects of the soul—the highest being reason, and the lowest being the passions or appetites. While it is true that overpowering emotion sometimes detracts from our freedom, there is increasing recognition that reason and emotion ordinarily work together. Daniel Dennett points our that apart from the "interests" attributable to organisms (survival, replication), there would be no reasons in the universe, only causes.

> How could reason ever find a foothold in a material, mechanical universe? In the beginning there were no reasons; there were only causes. Nothing had a purpose, nothing had so much as a function; there was no teleology in the world at all. The explanation for this is simple: there was nothing that had interests. But after millennia there happened to emerge simple replicators, and while *they* had no inkling of their interests, and perhaps properly speaking had no interests, we, peering back from our Godlike vantage point at their early days, can nonarbitrarily assign them certain interests—generated by their defining "interest" in self-replication.[61]

Similarly, Austin Farrer wrote in his Gifford Lectures that without "subjective factors", nothing coming before the mind would have the least tendency to produce action. "Even mortal danger would leave us unmoved if we had no repulsion against death or bodily harm."[62]

Antonio Damasio's research on patients with damage to the medial portion of their frontal lobes provides the beginning of an understanding in biological terms of how emotional processes provide signals about aspects

[60] Alasdair MacIntyre, *Dependent Rational Animals: Why Human Beings Need the Virtues* (Chicago: Open Court, 1999), ch. 4.

[61] Dennett, *Elbow Room*, 21.

[62] Austin Farrer, *The Freedom of the Will* (London: Adam & Charles Black, 1958), 127.

of our knowledge of the world that are not directly available to conscious awareness, and thus, how emotions are necessary for making reasoned judgments about potential action (see Ch. 1, sec. 3.2 above). Thus, reasons would not move us were it not for interest, affections, passions, and emotions.

8.1.2. Free Will as Autonomy The central meaning of "autonomy" is to be self-governed or self-controlled. Threats to autonomy can come from outside (i.e., other people) or from "inside". Autonomy in the first sense relates to the issues of political freedom, but our emphasis here is on the importance of autonomy from others in becoming what MacIntyre calls "independent practical reasoners".

> What each of us has to do, in order to develop our powers as independent reasoners, and so to flourish *qua* members of our species, is to make the transition from accepting what we are taught by those earliest teachers to making our own independent judgments about goods, judgments that we are able to justify rationally to ourselves and to others as furnishing us with good reasons for acting in this way rather than that.[63]

This transition involves coming to realize at a certain point that it will please parents and teachers *not* to act so as to please them but to act "so as to achieve what is good and best, whether this pleases them or not".[64] So we are dependent, in becoming autonomous moral reasoners, on teachers who have the moral abilities themselves to allow for independent judgment.

The second sense of autonomy, better captured by the term "self-control", is the opposite of Aristotle's *akrasia*. In the tradition there has been a tendency to equate self-control with the rule of reason over passion. Alfred Mele describes self-controlled people as those who "have significant motivation to conduct themselves as they judge best and a robust capacity to do what it takes so to conduct themselves in the face of (actual or anticipated) competing motivation".[65] Aristotle had focused his discussion on temperance, but Mele rightly broadens the sphere of self-control beyond temperance to include the emotions and also belief. With regard to the latter, Mele points out that one can be *akratic* in one's believing—for

[63] MacIntyre, *Dependent Rational Animals*, 71. [64] Ibid. 84.

[65] Alfred Mele, "Autonomy, Self-Control, and Weakness of Will", in Kane (ed.), *Oxford Handbook of Free Will*, 529–48, at p. 531.

example, allowing oneself to be self-deceived. He also objects to Aristotle's assumption that it is reason alone that is the arbiter of what one ought to do.

> On an alternative, holistic view of human beings, the "self" of self-control is identified with the whole person rather than with reason. Even when one's passions and emotions run counter to one's better judgment, they often are not plausibly seen as alien forces. A conception of self-controlled individuals as, roughly, people who characteristically are guided by their better judgments even in the face of strong competing motivation does not commit one to viewing emotion, passion, and the like as having no place in the "self" of self-control. Self-control can be exercised in support of better judgments partially based on a person's appetites or emotional commitments. In some cases, our better judgments may indicate our evaluative ranking of competing *emotions* or *appetites*.[66]

This view is compatible with our account in Chapter 6. Responsible *reflection* involves evaluation of desires and impulses, but also of beliefs and perceptions. Morally responsible *action* involves acting in accord with the higher-order judgment, but there is no presumption here that reason *per se* is a "higher" faculty; "higher" in our sense is related to levels of self-transcendence.

8.1.3. Hierarchical Mesh Theories of Freedom As noted above, Harry Frankfurt helpfully distinguishes between first- and second-order desires; second-order desires are desires about having or not having one's first-order desires. There is the common experience of desiring that we *not* have our own first-order desires, such as in the case of wanting to break a habit. Frankfurt argues that a person is responsible for his actions *and free* when his second-order desires are consistent with the lower level. Otherwise, he says, one is a passive victim of one's desires. So freedom consists in "motivational wholeness", regardless of the origin of one's desires.[67]

We endorse Frankfurt's recognition of the role of higher-order evaluation of one's desires, but we believe several modifications of his position are in order. First, as Charles Taylor has argued, we often act for desire-independent reasons. Taylor speaks of strong evaluation in contrast to weak evaluation, the latter based merely on something's being

[66] Alfred Mele, "Autonomy, Self-Control, and Weakness of Will", 532.

[67] Harry Frankfurt, "Alternative Possibilities and Moral Responsibility", *Journal of Philosophy*, 66 (1969), 829–89.

desirable. Strong evaluation invokes moral concepts.[68] Thus, we have adopted MacIntyre's account of moral responsibility in terms of evaluating "that which moves one to act"—he recognizes that desires are but one sort of motive—and whose evaluation is ultimately in terms of moral concepts.

Second, Frankfurt speaks of conforming one's higher-level desires to lower. We would emphasize, in contrast, the role of higher-order evaluation in allowing for the adjustment of lower-order desires (and other motives). Responsibility comes from having taken part in shaping one's own desires and attitudes. For example, inhibition of desired actions leads to habits, which are instantiated by means of neural changes. Such changes may result in making conscious inhibition unnecessary. We can also make changes in our environments now to forestall later desires and drives.[69]

8.1.4. Agent Causation There are a vast number of positions on free will that invoke the notion of agent causation.[70] We shall not attempt to survey them here. One motive for developing the concept of agent causation has been the perceived impossibility of reconciling free will with ordinary physical causation. Thus, a special category of causation has been postulated whereby agents initiate sequences of events without that initiation being itself caused or determined.[71] Note that this apparent requirement comes from assuming a simplistic linear model of causation in the human sphere, which we criticized above (sec. 5).

Another, more legitimate, reason for such developments is recognition of the difference between mechanistic and purposive explanations; that is, explaining why an agent did something is clearly different from explaining a mechanical process that brought about some effect. Our argument to the effect that ordinary causal processes, appropriately structured, enable purposive behavior is relevant here. However, we recognize purposive

[68] Charles Taylor, "What is Human Agency?", in *Philosophical Papers*, 1 (Cambridge: Cambridge University Press, 1985), 15–44. See also John R. Searle's account of the creation of desire-independent reasons, in *Rationality in Action* (Cambridge, Mass.: MIT Press), ch. 6.

[69] See Ch. 6, sec. 3.6, for our report on George Ainslee's *Breakdown of Will* (Cambridge: Cambridge University Press, 2001).

[70] For an overview, see Timothy O'Connor, "Libertarian Views: Dualist and Agent-Causal Theories", in Kane (ed.), *Oxford Handbook of Free Will*, 337–55.

[71] Simon Blackburn, "Agent Causation", in *idem* (ed.), *The Oxford Dictionary of Philosophy* (Oxford: Oxford University Press, 1994), 9.

behavior in non-human animals, so a two-category distinction between mechanistic and purposive is not adequate. There is a spectrum from purely mechanical causation through the purposive behavior of primitive organisms, the intelligent goal seeking of higher animals, and finally to the reason-guided teleological action of humans.

8.2 The Achievement of Free Will

In our account of the differences separating mature humans from inanimate matter we have sketched a variety of levels of complexity, exhibiting increasing capacities for flexibility and self-direction. Even primitive organisms exhibit goal direction and action under evaluation. So what are the differences between humans and lower organisms? Unlike the lowly protozoa with only two possible courses of action—to swim ahead or turn—we have an extremely wide repertoire of behavioral possibilities. Unlike *Sphex*, we have the ability to learn from our past mistakes. As with nearly all other animals we can change from the pursuit of one goal to another in light of feedback from the environment. So increasing abilities to sense the environment, coupled with increasingly complex behavioral repertoires, result, in higher organisms, in what we call acting reasonably, as when a dog attempts to open a cupboard in which it smells food.

The development of human culture has resulted in the creation of external scaffolding, which leads to a remarkable increase in rational abilities. Language allows us to respond to the world around us in terms of abstract categories. The development of theories of rationality and logic, along with a symbolic concept of the self, allows us to engage in meta-level evaluation of our cognitive strategies. We argued in the previous chapter that such meta-evaluation of one's reasons for acting constitutes moral responsibility.

So our minimalist model of a moral agent is represented as in Figure 3.4. This is a picture of a person in her environment who, in Dennett's terms has "gone meta" in order to make her own lower-level cognitions, propensities, behavior, *and criteria for evaluation* (MC in the diagram) the object of evaluation. Juarrero says that "intentional human action is free to the degree that contextual constraints put the most complex levels of its neurological organization, those governing meaning, values, and morals, in control. Even as these levels regulate and close off

options, they simultaneously free up qualitatively new possibilities for the expression of those values and morals.''[72] As noted in the previous chapter, this process of self-transcendence can continue with the adding of additional levels of supervisory systems. It is interesting to note the grammatical fact that one associates the ''I'', the true self, with the highest-level supervisory system. I (nominative case) evaluate my motives.

The concepts of free will and moral responsibility are so closely linked in the philosophical literature and in ordinary discourse that many authors (e.g., Frankfurt) use them interchangeably in the course of their arguments. So is it reasonable to conclude that the criteria for one are indeed identical to those for the other? We have shown that our morally responsible actor fulfills all of the other traditional criteria for free will (to the extent that one should want). The one who is able to evaluate that which moves her to act (Frankfurt's lower-order desires included), on the basis of reason (Kant), especially moral reasons (Taylor, MacIntyre); who is furthermore not unduly influenced by the judgments of others (autonomy in the first sense), nor prevented from acting according to that evaluation by weakness of will or overpowering emotions (Mele; autonomy in the second sense) is indeed an agent, the primary, top-down cause (agent causation) of her own actions. Our suggestion, then, is that free will be understood as being the primary cause of one's own actions; this is a holistic capacity of mature, self-reflective human organisms acting within suitable social contexts.

9 Conclusion: An Agenda for Future Research

Our goal in this book has been the defeat of neurobiological reductionism, which we have taken to include the denial of a physical organism's ability to be the cause of its own action, to use language meaningfully, to act for a reason, and in a morally responsible way. In this chapter we addressed one aspect of the free-will problem, the worry that human behavior is biologically determined. In fact, we believe that greater understanding of our own neural and cognitive processes sheds valuable light on the

[72] Juarrero, *Dynamics in Action*, 249.

concept of free will itself. We suggest, then, that the problem of reconciling free will with neurobiology is *not* a particularly knotty philosophical problem, but rather an agenda for research—namely, to provide increasingly well-informed and cogent accounts of how our neural equipment provides for the cognitive and behavioral capacities that go into responsible action.