

# 1 Introduction: Science Catches the Will

**Don Ross**

If there were an all-star list of concepts from the history of western philosophy based on the volume of attention over the years, the concept of will would be among the list's elements. The reasons for this are not obscure. It has been taken as a principal source of human specialness that we are putatively original authors of some events, which themselves then get dignified as a select subset called "actions." Exercises of will have been regarded as a sui generis type of process, events of "agency." This in turn opens a raft of questions as to who else besides the prototypes of agency, individual human beings, might also partake wholly or partly of it. Animals? Infants and cognitively impaired people? Suitably organized and structured groups of people? Integral parts or functionally distinctive parts—'faculties'—of people? These questions in turn seem to rest on resolution of others. Does agency essentially involve rationality, and if so, of what kind and beyond what threshold? How can the apparently close relationship between agency and motivation by reference to reasons be squared with the intuition that spontaneous, subjectively nondeliberative decisions to act seem like the paradigmatic exercises of will—even though someone who always and only acted spontaneously and unpredictably would likely not be deemed an agent at all, and so not be thought to properly will anything? During every age of western philosophy without exception these questions, and the assumptions required to make sense of them, have been in prominent play.

The concept of will is central in pre-modern Christian philosophy for obvious reasons: the main theme of the Christian myth is a gift of free will to humans by an omnipotent god, resulting in a drama that is complex, to put it mildly. The consequent entrenchment of a conceptual network based around the will in all aspects of western moral culture left the notion less vulnerable than most of its fellow pre-modern conceptual all-stars when the scientific revolution arrived. The majority of early modern philosophers were inclined to accommodate the concept of will into scientific metaphysics

rather than aim to displace it or explain it away. Hume, as usual, is the great exception, while his foil Descartes also plays his customary role as the exemplary synthesizer. Ingeniously, Descartes borrowed the Christian conception of will as the basis of both human specialness and human sinfulness so as to buttress his version of science-friendly epistemology. That is, will was presented by him as a necessary part of the source of all error, from which it required rescue and discipline by the faculty of understanding. But because he allowed the will to retain its role as the source of *sui generis* mental *activity*, he was able to appeal to it to try to insulate his program for sweeping mechanization of nature from both apparent counterexamples based on nonmechanical human agency and from dangerous conflicts with the prevailing moral and social order. There were too many sharp minds in the church of Descartes's day for this cunning to turn the intellectual authorities then and there, but despite losing almost all battles to his critics in the debates following the *Meditations*, he gradually—posthumously—won the war. The efforts of Hume and Nietzsche notwithstanding, Cartesian dualism became the basis for the prevailing popular metaphysics of both the natural order and of morality in western culture. So it clearly remains.

However, the academy—or at least that part that is led by science—has almost all defected from the popular picture. Among professional students of mind and behavior, dualism has few adherents. However, support for the idea of free will, under some alternative interpretation, probably still commands majority assent, at least if the alternative is taken to be the thesis that the appearance of agency is an illusion. Two simple propositions stand in the way of banishing the concept of will to the realm of caloric and demonic possession—and the Cartesian faculty of understanding. First, it is difficult to see how attributions of moral responsibility can be justified if no one is really the author of their actions. It is, of course, generally recognized among philosophers and psychologists that discomfort at a belief's consequences cannot be invoked to refute it. If we have been blaming and punishing people, on a vast scale, who in fact cannot possibly deserve it, then we should be prepared to see our moral culture and the institutions that rest on it revised. However, contemplation of this vertiginous challenge is discouraged by the second simple proposition: no argument against the existence of the will, however cogent, seems to carry conviction stronger than everyone's sense that they can, for example, decide to raise their left arm and then feel and watch it go up. And if we can each autonomously choose to raise or not raise our arms, presumably we

can autonomously choose to pilfer or not pilfer the pension fund and poison or not poison our enemies.

Thus, although dualism is a marginalized intellectual preference these days, the pivotal concept of will on which it has always rested remains philosophically current. We can note signs of unease in semantic trends, however. Philosophers and scientists tend no longer to use the word “will” by itself—that is, apart from its occurrence in the phrase “free will.” This reflects the fact that no one continues to believe in classical faculty psychology (popularity of cognitive modules in some quarters notwithstanding). Thus while free will might be the most natural name for a putative real phenomenon, the idea that it resides in the capacities of some sort of *entity* is disreputable. When philosophers are inquiring into the nature of the will itself, as opposed to free will, they are more likely to say they are investigating agency. I do not advise them against this more cautious language, but it can obscure the fact that in the worldviews of many avowedly naturalistic philosophers a traditional pre-modern concept still influentially lurks.

Among philosophers who have recently wrestled with this challenge, the two most widely read are Donald Davidson and Daniel Dennett. Davidson’s approach is analytic: he tries to break down the monolithic problem of how will can exist in a natural world into less grandiose constituent problems that, while still hard, might be tractable.<sup>1</sup> Such problems include “Can reasons be causes?” and “How is it possible that people sometimes act so as to disappoint themselves with respect to their own standing preferences?” If these questions and others like them can be given sensible naturalistic answers, then in the course of working them out we are in effect learning how to reconcile the familiar inherited conceptual network around the will with our scientific metaphysics. Dennett, by contrast, has preferred to tackle what I called the monolithic problem synthetically and head-on: in two books (1984, 2003) and many supporting articles, he has argued that much of what has traditionally been said about the will is indeed shown by science to be illusory, but that the consequences of this for everyday moral life have been greatly exaggerated. We have, as Dennett famously puts it in the earlier book, all the free will worth wanting. Endorsing this conclusion, if we do so, can still leave us wondering whether we have all the will, simpliciter, worth wanting from the point of view of a metaphysics of the person that is both comprehensible and in accord with scientific evidence; this is the issue that motivates Dennett’s second book on the subject.

The pressure of the scientific evidence in question has recently become much more acute than in the past. The main source of early modern tension, the fact that mechanical determinism was a broadly successful program in post-Newtonian physics, might have made the will into an object of philosophical suspicion, but it can hardly be said to have constituted true scientific evidence against it. However, the vast increase in the sophistication of the brain and behavioral sciences over the past few decades changes the situation dramatically. We are beginning to understand, on the basis of direct empirical investigation, how human behavior at various scales of analysis is controlled and influenced. From this perspective, the will has always mainly been a black box into which have been bundled all efferent behavioral control factors that have been inferred to exist but remained unexamined by science. As long as that set included *most* of the hypothesized factors, philosophers could freely speculate without inviting justified charges of fecklessness. This circumstance no longer prevails.

The factors that are lately being dragged out of the black box arise on both micro and macro scales. By “micro” factors I refer to influences on short-term decisions and fine-grained calibration of action: that is, within-brain causal antecedents of raising or not raising arms, pulling or not pulling triggers, and, in a realm of pressing policy relevance, taking or not taking another drink, cigarette, slice of pie, or pull on the slot machine lever. By “macro” factors I refer to influences on the gradual sculpting of personalities and selves, as manifest in dispositions to particular patterns of action, operating at time scales measured in months and years. I do not intend here to urge a binary distinction, but to draw attention to a continuum of scales by focusing on two contrasting points. My examples above do not capture the extrema on the continuum. For those, consider on the one side a baseball player “deciding” to tip his bat just up or just down as the pitch crosses the plate, which cannot possibly (because of processing speed considerations) be a *personal* decision in the sense of involving his deliberative consciousness or even his frontal cortex. On the other extreme, a human personality is partly constrained by patterns in natural selection that have unfolded over hundreds of millions of years. On both of these limiting sides, many think or feel that we have passed beyond the domain of will and into that of exogenous causation.

Moral culture is heavily preoccupied with aspects of agency, and with its limits, on all scales. We hold people responsible for pulling triggers and taking drinks. We massively reward good characters and try to blight the lives of bad ones. Controversies over the relative contributions of “nature” and “nurture” to human character are conducted tirelessly and with moral pas-

sion on both sides, and social scientists are often resented for promoting evidence in favor of structural rather than agent-driven causes of outcomes. Most societies allow insanity defenses to reduce or remove criminal liability, but this is controversial with conservatives. Millions of people around the world deny rationally incontrovertible evidence for evolution because they fear that natural selection threatens the sense of human and divine autonomy. Issues around the nature of will and agency thus embroil leading fronts of scientific progress directly in ideological, political, and legal tempests.

Nearer to the micro end of the continuum, science is opening the black box of the will in two main ways. First, ingenious behavioral experiments of a kind pioneered by Libet (e.g., 1985), and extended by subsequent researchers whose leading representative is Wegner (2002), have confounded basic widespread assumptions supposedly grounded in everyday unreflective experience to the effect that events such as conscious subjective choices to raise arms must temporally precede all arm-raising that are not brought about by extra-personal forces. The brain, it turns out, prepares such actions before its personal, conscious “operator” is aware that it is doing so. Furthermore false impressions of being consciously and subjectively in control of micro-scale actions as they unfold can reliably and relatively easily be induced by manipulation of environmental conditions. I will not try to further summarize this work here, since the reader is about to encounter extended descriptions of it in several of the chapters to come. Suffice to say that it seems on its face to undermine, or at the very least to greatly complicate, the second of what I identified above as the two “simple propositions” that have allowed the Cartesian will so much longer a life in serious inquiry than most of its conceptual kin.

The other avenue along which science is disrupting traditional ideas of will and agency at the micro scale is through attention to the architecture of the mind/brain. Descartes believed that the body, including what we now call the nervous system, was a mechanical system, basically a hydraulic network of pulleys and levers. The will’s task in administering this network was simply to increase and decrease distributions of tension at the site in the brain where chords and cables converged. This assumption contributed greatly to the plausibility of the *simplicity* and *unity* of the Cartesian will, from whence it derived most of its explanatory power. As vividly described by Dennett (1991, and elsewhere) and Glimcher (2003), Descartes’s conception of the will as a fused control point that coordinates and regulates reflexes in discharge of action plans it itself spontaneously originates not only survived through many decades of early neuroscience, it

positively shaped the dominant research paradigm in that discipline due to Sherrington.

Over the past fifteen to twenty years, this paradigm has been blasted to ruins. First, experience in artificial intelligence demonstrated that real-time management of complex behavior involving networks of interrelated sub-routines comes with stringent limitations on the extent to which sequencing of actions can be allowed to bottleneck at central control points. Early AI designs for performing even severely limited subsets of the everyday human task repertoire tended to resemble mini Soviet Unions, grievously incapable of supplying behavioral responses adequate to meet environmental demands the moment the latter were allowed to become the least bit difficult to predict and monitor (Brooks 1999). The nexus of brain and environment, it came generally to be recognized, is a complex system dominated by nonlinear feedback, amplification and damping, that can be (imperfectly) managed only by information-processing systems that are themselves complex (Thelen and Smith 1994; Port and van Gelder 1995).

Several philosophers of mind and cognitive science, including Dennett, have contributed significantly to developing the implications of complex, distributed models of cognition and control. Most would agree that the foremost figure among these philosophers is Andy Clark. After devoting his early work to issues surrounding the kind of cognitive architecture needed to account for manifest behavioral patterns (both competencies and characteristic breakdowns and errors), Clark turned to cultivating a truly radical consequence of the recognition that the will cannot be a simple lever or identified with a spatiotemporal point. This is that the will, insofar as it is identified with the efferent aspect of the agent, person, self or mind, cannot nonarbitrarily be contained within the brain and body of the human organism (Clark 1997, 2003; Clark and Chalmers 1998). I will again not attempt to summarize this perspective or the arguments for it here, since Clark himself attends to this task in a later chapter. The basic argument schema may be stated quite straightforwardly, however. When the Cartesian will is regarded as the nonphysical principle of activity, the whole brain is effectively made part of the external environment. But once we abandon dualism and begin distributing control of behavior, including such behavior as we deem fit to call "action," around in the brain, our hold on the boundaries of the will begins to slip, along two dimensions.

First, a point implicit in all distributed-control models, it becomes entirely unclear how to distinguish the will, now no longer a faculty that is among the mind's components, from the whole agent, which is in turn often treated as synonymous with the self and the person. Second, and

more distinctive of Clark's position, once we allow that "devices" within the brain, such as perceptual systems, can be aspects of the self because they exercise control functions (both directly and via their roles in complex feedback loops), it emerges as arbitrary if we try to necessarily consign external prostheses and environmental scaffolds to the will's external environment merely because they are not implemented in the body's cells. People, like many other animals, reorganize their environments in idiosyncratic ways to help regulate their own behavior. Is a typical person's inner ear less a part of herself than her computer, organized as it is to remind her of her projects, organize and conserve her latest thoughts and cue new ones that cohere with them, and manage a large part of her communication with others? Are you performing an action as an agent, responsive to reasons rather than causes, when you multiply two large numbers? You know you couldn't do this without your frontal cortex, and you're inclined to take your frontal cortex as part of the system that implements your will (or selfhood, agency, etc.). But could you multiply the numbers without the pen and paper or calculator you use? Could you do it without a culture that provided a symbolic notation for conceptualizing your task in the first place, and then for keeping track of the subroutines it requires? So if cortex is part of the system that implements your will—which implements *you*—why not your essential cultural technology for doing arithmetic?

If we are persuaded by Clark and like-minded thinkers to allow the self/person/mind to be distributed outside the organism casing, then the undermining of the traditional will from the micro direction meets and makes common cause with such undermining from the macro side. As the examples above are intended to illustrate, the exploding new interdisciplinary sciences (evolutionary psychology and anthropology, behavioral and institutional economics, multi-agent modeling, complex system simulation) that study cultural evolution, imitation, convergence to equilibria in games, and the unplanned origins of institutions and norms, all emphasize the extent to which interpersonal dynamics establish the enabling conditions for the actions of persons qua relatively behaviorally unified agents. But then, and especially in conjunction with micro-scale influences from subpersonal processes in the brain, they seem to make will and agency redundant.

The work of Libet, Wegner, and Dennett suggests that your brain is built to be able to initiate and steer many of your actions with little or no deliberation or phenomenal awareness. Then consideration of social dynamics suggests that a necessary condition for your brain's ability to stay on track is a relatively stable social environment, furnishing norms and targets for

imitation that have evolved in the culture to get precisely such jobs done. At this point there seems no evident task left for a vestigial will. Just as well, one might think—another bit of prescientific metaphysical litter goes into the bin.

But now recall that when we put pressure on the will as a Cartesian point-mass it became increasingly difficult to distinguish from the concept of the self. If the will is eliminated, does not the self disappear with it, and for the same basic reasons? Can the progress of science convince us that there are no individual persons but only biological individuals pushed into certain sorts of behavioral dispositions rather than others by the complex dynamics interlocking their brains and their cultural/social environments? Note that in wondering about the existence of persons without calling into question the existence of distinguishable instances of *H. sapiens* we are not raising doubts about the status of any biochemical or biopsychological facts. Some might take us to be wondering about a metaphysical fact, but this does not interest scientists, and even many philosophers deny that there are such facts. If the question seems important, it is because persons—as opposed to human organisms—are at the center of moral culture. It is indeed difficult to imagine how people would function, would indeed even be people, if they did not suppose that they were substantially responsible for at least their own moral and social choices in adolescence and adulthood. The concepts of the will and the self appear to rise or fall together.

Among contributions to this volume, I have explicitly mentioned the work of Dennett, Wegner, and Clark as leading elements of the book's intellectual background. One further scene-setter now needs introduction. I noted earlier that although talk of "free will" is still common at least in philosophy, the "will" simpliciter is rarely invoked due to the conviction that it is not a mental organ, and that there's no evident *other* sort of entity it could be instead. Recently, however, the psychiatrist and behavioral economist George Ainslie has rehabilitated frank talk of the will. He has been able to do this because he has found a new ontological interpretation for it. According to Ainslie, the will is not an organ but a *habit*, a recurrent (to considerably varying extents from person to person) pattern in people's behavior. Development of this perspective begins from observing that people's (and other animals') preferences are sensitive to anticipated delays of rewards in a way that makes them singularly ill-suited to maximization. In particular, they are discounted into the future by hyperbolic-shaped functions that lead preferences to temporarily reverse as an individual's tempo-



ral distance to the elements in her reward stream changes (see Ainslie's chapter for details). This goes a long way, Ainslie argues, toward explaining addiction and other pathologies of rational choice identified by behavioral economists. But then it raises a puzzle as to how most of us rise above our natural inconsistency so as to approximate the prudent investors recognized as agents in microeconomic theory. Ainslie proposes that the answer to this puzzle lies in the ability to see that inconsistency now *predicts* defeat of *currently* entertained long-term projects. As a result disappointment over future prospects can arise not just in the future, about which discounting makes us unduly cavalier, but in the more highly motivating present. Furthermore the tendency to pay attention to the predictive aspect of preference consistency and inconsistency is something a person can cultivate and become habituated to. Such cultivation and habituation amounts, says Ainslie, to creation and maintenance of will. So here we have a perspective in which the will is in no way metaphysically mysterious but is also cut loose from its conceptual background in Cartesianism and faculty psychology. It is particularly interesting in this connection that Ainslie further models the will as emerging from the bargaining activity of subpersonal interests—something that follows, he argues, from hyperbolic discounting—thus echoing the general theme of distributed cognition theorists. Simultaneously, his picture depends on the idea that what induces the bickering and selfish subpersonal units to settle their differences and, in so doing, give rise to willpower are social pressures: if your gang of interests fails to coordinate while your neighbor's coalition (forced together like yours through inhabiting the same agent at the macro scale) succeed, then yours will all jointly stand to be exploited by the more consistent team. So, far from taking the concept to be *rejected* by new theories of distributed cognition and social-dynamical influence on stabilization of selves and characters, Ainslie offers a new model of will that revives the idea by appeal to those very resources.

Thus inspired by the possibility that the apparent threats to the cogency of the idea of will might instead turn out be the basis of its reemergence as a scientific subject, we gathered together the contributors represented in this volume to evaluate the prospects. The banner for our deliberations was the Mind and World (mAw) working group founded at the University of KwaZulu-Natal in eThekweni (Durban), South Africa, in 2002. mAw's aim is to develop and apply models of mind as distributed and extended. It has built a close working relationship with the Center for Ethics and Values in the Sciences at the University of Alabama at Birmingham. The

mAw conference on Distributed Cognition and the Will, co-sponsored by UKZN and the Center, was held in Birmingham on March 18–20, 2005. The papers gathered as chapters in this volume are descendants of the presentations given there, refined by virtue of the extended critical discussion they individually and collectively provoked.

The chapters address the status of the concepts of the will and the self on the basis of first taking for granted that the recent scientific developments that put pressure on them must not be denied, avoided, or lazily interpreted so as to seem harmless to traditional views. This volume thus does not represent a debate between naturalists and others; all contributors are motivated by a commitment to take science seriously. Only about half of the chapters are by philosophers; the remaining half are by behavioral scientists of various disciplinary persuasions.

I will briefly survey the contents to come in order to give the reader an idea of the thematic trajectory we had in mind in choosing an order for them.

Chapter 2 by Daniel Wegner and Betsy Sparrow summarizes the most basic source of the new scientific challenge to the traditional concept of the will: the empirical evidence, gathered to a significant extent in Wegner's lab, that our standard perception of the subjective experience of will as causing our voluntary behavior on micro scales is an illusion. Wegner and Sparrow then propose an explanation for this illusion that introduces the book's other main themes. The explanation in question is that the illusion of will has been selected (both biologically and culturally) because in underwriting our sense of responsibility for our effects on social stability, it promotes maintenance of that stability, from which individuals in turn derive benefits. Thus the relationships among neural dynamics, behavioral dynamics, and social dynamics as the basis for the phenomenon of the self as a stabilization device, are put onto the table at the outset. They remain there throughout the volume.

In chapter 3 Paul Sheldon Davies raises reasons for fearing that Wegner and Sparrow are too quick to seek an ecumenical accommodation of our traditional cultural model of the agent in the face of their data. This is a recurrent issue for naturalistic philosophers: to what extent should we aim to domesticate the counterintuitive findings of science so as to go on feeling comfortable in our conceptual skins, and to what extent should we embrace the exposure of our folk myths as we learn more about a universe that was not designed, and in particular was not designed to be understood by us (since natural selection didn't directly favor those of our ancestors who could explain any but a handful of ecologically local phenomena)?

Davies's chapter 4 leads straightforwardly into Tamler Sommers's application of the same general concern to the more specific question of whether our scientific knowledge is compatible with the idea that people as moral agents are causally responsible for their actions. Sommers doesn't just wonder about this: he argues forcefully for incompatibility, for what might be called cold-water naturalism about moral responsibility understood on the model of causal responsibility. Though defending this conclusion is Sommers's main purpose in the chapter, he follows the general line of Dennett's 1984 book on free will—but *not* the line of Dennett (2003). He suggests that if we take his advice and give up our illusion of responsibility, this need not have the corrosive effects on social stability that the (conceptual) conservative fears, and that might be thought to follow from Wegner and Sparrow's explanation of the illusion of will.

Phillip Pettit, in chapter 5, provides a much more sustained argument for Sommers's second conclusion above. Pettit has long been among the foremost developers of the theory of agency as a social phenomenon. He extends that development in showing how the will need not cause any actions on micro scales in order for our responsibility as agents to be truthfully attributed and to be functionally selected by cultural evolution so as to regulate behavior that underwrites social stability. Agency, according to Pettit, merely requires our conforming our socially significant behavioral dispositions to answerability to reason on relatively macro scales, and this in no way requires control of fine-grained actions by the will on micro scales. It might be said that this idea emerges as nearly a consensus in the book. Of all the authors here, only Davies seems to be somewhat skeptical on the point.

The chapters to this point, after Wegner's and Sparrow's survey of evidence in the first half of theirs, all concentrate on the implications for moral and social agency of skepticism about willful causation on the micro scale. In chapter 6 Daniel Dennett addresses this too, but also begins to widen the target to consider the implications of recent behavioral and cognitive science for our ability to understand ourselves not just as moral agents but as agents or selves *at all*. He is not in the least skeptical about this prospect, but he thinks we have hard work to do in order to recover a clear picture. Like most authors in the book, Dennett thinks that our micro-scale agency is based on a foundation of macro-scale agency. He agrees with Pettit in emphasizing the crucial role of interpersonal communicative coordination. But whereas Pettit emphasizes coordination on standards of reason for the sake of justification of action, Dennett draws attention to a logically prior level on which we are under pressure to make agents of

ourselves: if we didn't, we couldn't engage in complex communication in the first place, and such communication is basic to the construction of the distinctive ecological niche our species has filled.

Having Andy Clark's chapter 7 on selfhood as an aspect of niche construction follow Dennett's allows us to recapitulate the wider history of the extended and distributed mind hypothesis. Dennett originally suggested and inspired it, but Clark fully developed and elaborated it. In his chapter Clark explains how selves exercise will (at appropriate scales) by means of what he calls "soft" or "ecological" control. Selves, he argues and explains, are problem-solving assemblages of expropriated ecological resources. Crucial among these resources, we might suppose, are culturally evolved standards of publicly sanctioned reasons for action. If this supposition is added to Clark's account, then it and Pettit's can naturally be conjoined with Dennett's to form a model of the self and its control capacities that replaces the one demolished by the evidence of Wegner and colleagues.

By this point in the book, more traditional philosophers might be grumbling that their patient analytical labors over the years are being casually brushed aside here by scientific show-and-tell. Where, they might ask, are the careful arguments required for systematically showing, in detail, that the distributed, extended, virtual, nonfacultative self and will can in fact perform the conceptually unifying function that motivated the old-fashioned versions of the concepts? Mariam Thalos's chapter 8 is the longest in the book because providing this patient engagement is the task she takes on. Agency, she shows analytically, must be distributed if philosophers are to have any hope of backing out of various logical cul-de-sacs into which they have driven themselves in their attempts to come to grips with the concept, and if they are to grasp the possibility of a scientific account of it. While nonphilosophers might find themselves skimming this chapter, readers and critics who are analytic philosophers are likely to regard it as the serious core of the book.

Philosophers will join everyone else in appreciating George Ainslie's chapter 9. It does two main things. First, it provides a concise introduction to his theory of the will based on hyperbolic discounting as sketched above, and his theory was the single most important motivator for the questions asked at our conference and in this book. This part of the chapter will be a valuable resource for readers new to Ainslie's model. Second, the chapter defends the value of thought experiments, often employed by Ainslie in developing his ideas, as guides to the nature of the mind and the person. Ainslie's thought experiments will intrigue philosophers because

most will be aware that the kind of naturalism promoted by Dennett, which forms another foundation block for the account of self and will in the book, has promoted skepticism about the possibility for thought experiments to contribute to empirical knowledge; Dennett has said on other occasions that most philosophers' thought experiments are failures of imagination disguised as insights into necessity. In this context it is surely interesting to find the importance of *some* thought experiments being defended by a scientist. Ainslie does some useful philosophy himself here by analyzing some of the features that distinguish illuminating thought experiments from those that indeed obfuscate in the ways Dennett has cataloged. If the reader accepts Ainslie's conclusion, it will be mainly because the sample thought experiments he provides indeed seem irresistible. To the extent that they are, the reader will discover that while her attention was officially on thought experiments, she imbibed compelling reasons to also endorse Ainslie's theory of the will.

Among those who agree that Ainslie's is the most promising and exciting positive model of the will now on offer, it might be seen as surprising that it comes from the discipline of behavioral economics. If any discipline has over its history treated the will as a pure black box, it is economics, and as Davis (2003) cogently argues, its leading neoclassical incarnation has not incorporated even a substantial concept of the individual person, let alone the self. However, while economists have indeed tried to get by with a *psychologically* thin conception of the person, they have made the concept of the *agent* fundamental to their science, and have done more careful modeling of agency in a vast range of real and hypothetical scenarios than any other inquirers. In chapter 10, I sketch a framework by which selves can be endogenously generated in a class of dynamic game-theoretic models that break no foundational rules of neoclassicism (thus asking us to surrender none of its axiomatic power). My approach specifically exemplifies the general theme of many papers in the book (including Wegner's, Pettit's, Dennett's, Clark's, Thalos's, and Ainslie's): the self is depicted as virtually created in niche construction, in order to perform the function of *simultaneously* stabilizing and intermediating the micro-scale dynamics of the distributed individual mind/brain and the macro-scale dynamics of society and culture. Selves, in my view, do not exist *despite* the complexity of these dynamics at both scales, they exist *because* of it.

The three following chapters lend symmetry to the book's organization. Early chapters (Sommers, Pettit, Dennett, Clark) concentrate on macro-scale issues. Middle papers (Thalos, Ainslie, Ross) work on the seam between macro- and micro-scale phenomena. The final three chapters focus

on the micro side and go deeper into the mind/brain. In chapter 11 Lawrence Lengbeyer provides a schematic model of how the distributed will might actually go about allocating its attention and control functions. Though his account is couched in conceptual terms and doesn't attempt to specify measurable parameters, he is careful to attend to its motivation in phenomena from experimental psychology.

A natural question the reader can carry forward from Lengbeyer's chapter to the following one, chapter 12, by Wayne Christensen is: To what extent is Lengbeyer's schematic model at the mental/functional level consistent with Christensen's detailed survey of what we know about the distribution of control in the vertebrate brain? Christensen sounds a strong warning that, as with most new and exciting ideas, enthusiasm for distributed control in cognitive science can carry theorists farther than the evidence warrants, and indeed to extremes where evident falsehoods are embraced. Christensen's sketch of the architecture of neural control is compatible with a good deal of *functional* distribution and decentralization, as he makes clear. However, he makes equally clear that there is a straightforward sense in which the brain incorporates a control hierarchy with an executive, subordinates, and a vector of information flow that makes attribution of these roles nonmetaphorical. After considering both Lengbeyer's and Christensen's chapters, readers might naturally want to ask: Does Lengbeyer's pmanager system have the sorts of attention-allocating and plan-guiding capacities to be the right functional specification of the executive neural system characterized by Christensen?

Christensen's sketch of neural control architecture is broadly cybernetic, in the rigorous sense of Wiener (1948). It can thus be taken as a solid empirical platform for Vancouver and Zawidzki's defense in chapter 13 of cybernetic control theory in applied psychology as against a dominant stream among practitioner advisors who seem to believe that modeling people as agents requires that we reject cognitive science. Vancouver and Zawidzki provide no evidence that their foils are familiar with Wegner's work; one surmises that if they were, they would be deeply disconcerted by the evidence, while at the same time confirmed in their view of the cognitive and behavioral sciences as dehumanizing. I hope that the chapters in this book—Vancouver's and Zawidzki's specifically, on one level, but then all the chapters together, on another level—provide compelling evidence that such a view is simply ignorant. Real science does not imply a threat to our status as agents answerable to social demands for justification, for the reasons explored by Sommers, Pettit, Dennett, Clark, Thalos, and Ross. It *does* imply a challenge to lazy "morality" that simply follows conven-

tion, endorses familiar platitudes and thereby invites sleepy quietism in the face of problems.

The book closes with a riveting application of this general perspective, in chapter 14, to a genuinely tragic problem, schizophrenia. Dan Lloyd first makes clear how devastating this widespread condition is, how resistant it is to effective intervention, and how baffling it must seem if one insists on cleaving to conventional models of agency and its relationships to reason and to the unity of the self. He then sketches an original model, supported by simulation evidence of his own and some clinical support, according to which schizophrenia is the symptom of dynamic coordination failure among parts of the distributed mind with respect to self-monitoring of attention and action cueing against time. This micro-scale breakdown pattern in dynamics is then suggested by Lloyd to be potentially analogous to loss of a democratic society's ability to function as a coherent agent—that is, as a genuine community—when standards of answerability to evidence in public discourse are debased, especially by the very officials most clearly responsible for maintenance of them.

Having followed Lengbeyer, Christensen, and Vancouver and Zawidzki into micro-scale manifestations of the distributedness of agency, Lloyd thus ends the book by returning the reader to its opening focus on the relationship between social stability and answerability to reason that, far from being overthrown along with the simple Cartesian will, is given increased emphasis when we buttress philosophical speculation with science and take the consequences seriously. Despite the different and occasionally conflicting perspectives on display in this book, we find here a firm consensus that we are collectively and individually better off for acknowledging that the comfortable conceptual image of the will and the self that has been characteristic of western culture for several centuries is paying diminishing returns, and for aiming to marshal our species' talent for coordinated collective agency to build a more accurate one.

#### Note

1. Davidson's most important work on the will (agency) is gathered in his 1980 book.

#### References

Brooks, R. 1999. *Cambrian Intelligence: The Early History of the New AI*. Cambridge: MIT Press.

- Clark, A. 1997. *Being There*. Cambridge: MIT Press.
- Clark, A. 2003. *Natural-Born Cyborgs*. Oxford: Oxford University Press.
- Clark, A., and D. Chalmers. 1998. The extended mind. *Analysis* 58: 7–19.
- Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.
- Davis, J. 2003. *The Theory of the Individual in Economics*. London: Routledge.
- Dennett, D. 1984. *Elbow Room*. Cambridge: MIT Press.
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little Brown.
- Dennett, D. 2003. *Freedom Evolves*. New York: Viking.
- Glimcher, P. 2003. *Decisions, Uncertainty and the Brain*. Cambridge: MIT Press.
- Libet, B. 1985. Unconscious cerebral initiative and the role of the conscious will in voluntary action. *Behavioral and Brain Sciences* 8: 529–66.
- Port, R., and T. van Gelder, eds. 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge: MIT Press.
- Thelen, E., and L. Smith. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: MIT Press.
- Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Wiener, N. 1948. *Cybernetics*. New York: Wiley.