

The Illusion of Conscious Will by Daniel M. Wegner. Cambridge, MA: Bradford Books, The MIT Press, 2002. xi + 405 pp., \$34.95 (cloth). ISBN 0-262-23222-7.

Many readers will likely find deliciously ironic, as did I, the idea that anyone could write a 400-page treatise with this title. But Daniel Wegner, Professor of Psychology at Harvard, is entirely serious. His first chapter sets forth clearly the central ideas and strategy of the book. His overarching goal is to reconcile two seemingly incompatible approaches to explaining human action: The first, which he regards as the only scientifically respectable approach, models itself after Laplace's infinite geometer, and aspires to explain every aspect of human behavior deterministically, as a strictly causal resultant of biology and prior experience (p. 1). The second, essentially useless for scientific purposes, is the everyday folk psychology which seeks instead to interpret and explain behavior as the conscious pursuit of courses of action which for the most part are chosen voluntarily. Wegner seeks to honor the importance of "conscious will" in the sense of a subjective experience that we cause our own behavior, while denying that it has in fact any causal potency. He acknowledges that most of the time our experiences and their causation appear to agree, as expected on the everyday view. However, this is not always the case, and the fact that they can sometimes disagree suggests that they may in fact arise from different sources. In the remainder of the book, therefore, he seeks to develop and support this thesis by systematically cataloguing situations in which the actual causation of behavior and the experience of having caused it seem to diverge. In general terms these are (1) situations in which persons believe that they have caused certain behaviors to happen even though they could not have done so, and (2) situations in which persons produce purposive-looking behaviors while denying any sense of having authored them.

The book is heavily researched and documented, and covers a large number of interesting topics, many of which are both complex and controversial. A short review cannot possibly do justice to this richness, but I will attempt to sketch the highlights.

Chapter 2, "Brain and Body", pulls together the small amount of existing neuroscience research directly bearing on the issues. Brain stimulation in various forms can produce actions with or without accompanying experiences of voluntariness, suggesting that these systems are at least partially separable. A few electrophysiological studies of elementary motor actions further suggest that the processes that actually cause the movement may occur prior to the subject's conscious decision to move. There is also much evidence showing that conscious processes are too slow to manage all the details of skilled and rapid action. Wegner interprets all such facts as supporting the idea that behavior in general is efficiently caused by unconscious brain mechanisms, while the experience of conscious volition operates on a different time scale, and is generated by different mechanisms.

Chapter 3, “The Experience of Will”, makes this view more explicit. A crucial figure (p. 68) claims that any action is directly caused by unconscious mental events, while an apparent causal path linking thought and action is fabricated by separate processes into the experience of conscious will. Our actions just happen to us, but thoughts of those actions are thrust into consciousness as previews of what we will do by additional unconscious mental events that may or may not be directly linked to those that actually cause the action. If our actions correspond sufficiently well to our thoughts about them we infer, retrospectively and incorrectly, that our thoughts in fact caused those actions. Good correspondence occurs when thoughts consistent with the action occur just prior to it, and no other good causal candidate is in sight. The balance of the chapter elaborates this “theory of apparent mental causation” (of which Wegner himself is the primary architect) and applies it in a preliminary way to various experimental and clinical phenomena including hallucinations and creative insight in which persons’ sense of authorship of their behavior is impaired.

Chapter 4, “An Analysis of Automatism”, applies the same theory to a variety of classic automatism such as table-turning, automatic writing including communication via Ouija boards or planchettes, dowsing, and ideomotor action. Wegner’s assumption throughout this whole exercise is clearly that nothing more is ever involved than ordinary “voluntary” actions performed by persons under conditions which somehow bypass or disable the mechanisms that would normally produce the correlated experience of conscious will. In this light automatism becomes the rule, and the experience of conscious will the exception (p. 143)—a secondary attachment to the more basic unconscious system that automatically produces all of our seemingly voluntary behavior.

Chapter 5, “Protecting the Illusion”, examines and undermines the human tendency to perceive ourselves as ideal causal agents in full control of our actions. This drive is so powerful that it can lead us to take responsibility for actions we perform even when we couldn’t have intended them, as for example in the case of post-hypnotic suggestions. Our concept of ourselves as causal agents is not fixed but has a developmental aspect; children only learn to explain their own behavior to themselves in terms of their conscious intentions over a period of time, and with difficulty. In fact, there is always a certain indeterminacy of action, because any complex action can be described in many ways, and hence “undertaken under one identity and later recognized under others” (p. 159). This affords flexibility in the interpretation of behavior in terms of intentions, and creates opportunities for confabulation of intentions. Our behavior just happens to us, but we adjust our perceptions and reports to preserve the illusion that we are causally responsible for it. In support of this picture Wegner reviews the sizeable social-psychology literature on cognitive dissonance, as well as neuroscientific studies with split-brain patients in which the dominant left hemisphere makes up stories to explain what the mute left hand is doing.

Chapter 6, “Action Projection”, explores conditions under which we can become convinced that actions we are performing ourselves are actually being

performed by someone or something else. Starting with the autokinetic effect, in which a stationary point light-source in a darkened room appears to be moving, Wegner passes via “clever” animals and “facilitated communication” to a discussion of the difficulties of determining “whodunit” (p. 208) in ordinary social interactions. The emphasis is again on application of the theory of apparent mental causation to understanding where and how such effects occur.

Chapter 7, “Virtual Agency”, deals with much more extreme cases of action projection, in which persons appear to invent entirely fictitious sources of their own actions. Beginning with ordinary sorts of imagination and role-playing, and a discussion of features that normally help us distinguish between the imaginary and the real, Wegner turns to his main examples—channeling and mediumship, spirit possession, and dissociative identity disorder (DID). He justifies this “walk on the wild side” (p. 224) in terms of its potential for shedding light on the nature of personal identity, correctly emphasizing the widespread distribution of the phenomena across history and across human societies. Seeking a middle path between theories that dismiss all such phenomena as role-playing or outright fakery and theories that seek to explain them all as manifestations of “trance” or altered states of consciousness, he advances the alternative view that they result from abnormal operation of the same processes that normally fabricate our illusion that we ourselves are conscious agents. DID patients are likened to computers that can run two or more operating systems, rather than the customary one. We are all robots, but most of us harbor only a single virtual agent.

Chapter 8, “Hypnosis and Will”, explores additional circumstances in which the experience of will and the force of will part company. While seeming to transfer their experience of control to another agent (the hypnotist), highly susceptible subjects also sometimes acquire unusual capacities for physiological self-control. Wegner skillfully explains the transfer of experienced control to the hypnotist in terms of the various elements of social influence and sequential acquiescence that go into the typical hypnotic setting and hypnotic induction procedures. He has considerably less success explaining the emergence of unusual abilities such as pain control and removal of warts, which he suggests may result simply from getting the normal conscious monitoring out of the way, so that automatic processes can function more uninterruptedly. He again seeks a theoretical position between “faking” and “trance” theories of hypnosis, while noting the recent appearance of functional neuroimaging studies that suggest the presence of real physiological changes accompanying deep hypnosis.

The final chapter, “The Mind’s Compass”, seeks to explain the role of the experience of will in our mental life. If it does not reflect an actual causative influence on our actions, what good is it? Wegner argues that the experience of conscious will is sufficient by itself to support functions normally thought to require the idea of free will—functions such as the sense of achievement and acceptance of moral responsibility for our actions. In his view the experience of

conscious will is simply a kind of cognitive emotion that marks certain of our actions, all of which are produced automatically by our organisms, as having been authored by us.

I have tried hard to be fair and objective in summarizing the content of this book, but it is probably already apparent that in contrast with the four cognitive scientists whose glowing tributes appear on the dust jacket (Christopher Frith, Bernard Baars, Gordon Bower, and Michael Gazzaniga) I find its central arguments anything but compelling. It is certainly true that the mind is not “transparent”, or fully accessible to introspection. This much is accepted by everyone nowadays, and Wegner’s book certainly performs the service of assembling in one place and in highly readable fashion a lot of material relevant to contemporary discussions regarding the content and organization of the “cognitive unconscious”. Beyond this, however, my reactions are almost entirely negative. It would take another volume of at least equal size to explain this in detail, but I will focus here on three main areas.

The central thrust of the book is an attempt to elevate Wegner’s theory of apparent mental causation, which arises primarily from experiments carried out with unselected undergraduate subjects under conditions deliberately contrived to be ambiguous in their causal structure in ways that normal experience is not, into a general theory of human action. I think this attempt fails completely. The fact that some actions or aspects of actions are automatic and not subject to voluntary control does not entail his central thesis that none are. What begin as suggestions to this effect early in the book undergo by constant repetition a kind of “hardening of the categories”, which gradually bestows upon it the status of established fact. Thus on the final page of text he says “Our sense of being a conscious agent comes at the cost of being technically wrong all the time”. But this is not established by the data presented, and in my view it is not even required on his own terms. He is perfectly willing to allow behavior to be caused by “unconscious mental events” or “unconscious thoughts” (whatever those are), so why not conscious ones? Wegner’s determination to evaporate the causal potency of conscious thought may derive in part from his association with philosopher Daniel Dennett, who pursues similar themes, and who is described in the preface as having “provided valuable commentary on the whole book”. But such views are regarded as unnecessarily extreme even by many committed physicalists. John Searle (1992), for example, openly aspires to obtain an account of the mind that is based entirely on the biology of the brain but can accommodate conscious mental causation of behavior. Like Searle, I regard this attempt to do away with the conscious mind as a causal agent as essentially a *reductio ad absurdum* of Wegner’s theory, which is pure epiphenomenalism virtually indistinguishable from that advanced by T. H. Huxley and other early physicalists.

Secondly, Wegner does not do full empirical justice to some of the topics he surveys. He is certainly to be commended for undertaking his “walk on the wild side”, because that is more than most contemporary cognitive scientists would

be willing to do, but he clearly does so with occasional nervous backward glances toward more conservative academic colleagues. The result is that the “walk” is carefully circumscribed in such a way as to prevent contact with additional kinds of phenomena that would gravely threaten the theoretical framework he is building. In particular he manages to avoid even mentioning the well-documented association between reports of paranormal capacities and events with mediumship, trance and possession, and hypnotism (for example, Gauld [1984,1992] and Kelly and Locke [1981]). He refers approvingly to Faraday’s superficial investigation of table-turning, but does not mention Crookes’s (1874) investigations with D. D. Home. He talks at length about automatic writing without ever mentioning works by contemporaries such as Ian Stevenson (1978), or Alan Gauld (op.cit.), which provide fuller and more balanced accounts of the relevant phenomena. His discussion of mediumship dwells heavily on its contemporary and relatively trivial expressions in “channeling”, and reflects at best scanty acquaintance with the classical literature of that subject (e.g., Broad [1962], Gauld [1984], Hodgson [1898], and Sidgwick [1915]). The same systematic constriction of his intellectual horizon occurs in connection with additional phenomena which though not strictly paranormal reveal the emergence in some cases of secondary personality of knowledge and skills that greatly exceed the capacities of the primary personality. Thus we get over a page about Bridey Murphy, but only a footnote about the far more interesting case of Patience Worth. The most critical phenomena, in short, are routinely ignored or trivialized in service of the theory.

This leads directly to my last and most fundamental criticism. Because of its unwillingness to confront the full complexity of its subject matter, this book in my opinion misses the opportunity to reach toward a much deeper understanding of concepts such as “person”, “self”, and “causal agent”. To do this one would have to reopen a long-neglected chapter in the history of psychology, one populated by momentous figures such as William James, Pierre Janet, and F. W. H. Myers. Wegner almost starts down this road; on p. 142 he quotes the following comment of William James regarding automatism, which goes to the heart of the subject: “The great *theoretic* interest of these automatic performances, whether speech or writing, consists in the questions they awaken as to the boundaries of our individuality. One of their most constant peculiarities is that the writing and speech announce themselves as from a personality other than the natural one of the writer, and often convince *him*, at any rate, that his organs are played upon by someone not himself” (James, 1889, p. 555). Wegner clearly does not take this possibility seriously—for him it appears axiomatic that no other real agent can be there (e.g., p. 224). In effect he simply takes over and restates in modern form the views of Huxley and Carpenter on “unconscious cerebration”. But these views were explicitly rejected by Myers, Janet, James and others of that period precisely because they failed to account satisfactorily for the fully developed empirical phenomena of automatism and secondary personality. Wegner’s treatment fails almost completely to represent that side of

the issues. His description of Janet's concept of dissociation as "the idea that the conscious executive part of the mind can become cut off from unconscious parts of the mind that produce behavior" is pure caricature. Janet and James both got their key ideas from Myers, who developed more fully than anyone else of that period an altogether different picture of the organization of the psyche in which multiple conscious causal agents can exist concurrently and with differing patterns of mutual awareness in conjunction with the same physical organism. Myers's fundamental idea of a something more in us that is of the same basic kind but represents the deeper part of our nature was carried forward by James throughout his later work, in particular *The Varieties of Religious Experience* and *A Pluralistic Universe*. Surely the most mature reflections of persons of this stature on the problems at issue cannot simply be ignored by a truly scientific psychology. I will not attempt to carry this subject further here, but interested readers are encouraged to consult Braude (1991), Crabtree (1993, 2003), and Ellenberger (1970), in addition to the sources already cited, for modern discussions that in my opinion do a far better job of coming to grips with the complex empirical realities of these phenomena.

EDWARD F. KELLY

Division of Personality Studies
Department of Psychiatric Medicine
University of Virginia
Charlottesville, VA 22908

References

- Braude, S. (1991). *First Person Plural: Multiple Personality and the Philosophy of Mind*. London: Routledge.
- Broad, C. D. (1962). *Lectures on Psychological Research*. London: Routledge & Kegan Paul.
- Crabtree, A. (1993). *From Mesmer to Freud: Magnetic Sleep and the Roots of Psychological Healing*. New Haven: Yale University Press.
- Crabtree, A. (2003). "Automatism" and the emergence of dynamic psychiatry. *Journal of the History of the Behavioral Sciences*, in press.
- Crookes, W. (1874). *Researches in the Phenomena of Spiritualism*. London: J. Burns.
- Ellenberger, H. (1970). *The Discovery of the Unconscious*. New York: Basic Books.
- Gauld, A. (1984). *Mediumship and Survival*. London: Granada Publishing.
- Gauld, A. (1992). *A History of Hypnotism*. Cambridge: Cambridge University Press.
- Hodgson, R. (1898). A further record of observations of certain phenomena of trance. *Proceedings of the Society for Psychical Research*, 13, 284–582.
- James, W. (1889). Notes on automatic writing. *Proceedings of the American Society for Psychical Research, 1 (old series)*, 548–565.
- Kelly, E. F., Locke, R. G. (1981). *Altered States of Consciousness and PSI: An Historical Survey and Research Prospectus*. New York: Parapsychology Foundation.
- Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Sidgwick, E. M. (1915). A contribution to the study of the psychology of Mrs. Piper's trance phenomena. *Proceedings of the Society for Psychical Research*, 28, 1–657.
- Stevenson, I. (1978). Some comments on automatic writing. *Journal of the American Society for Psychical Research*, 72, 315–332.