

Reviews

Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence

ANDY CLARK

Oxford, Oxford University Press, 2003

229 pages, ISBN: 0195148665 (hbk); \$26.00

Andy Clark is perhaps best known for his defense of an embodied philosophy of mind: cognition—and the mind more broadly—cannot be properly understood without a deep recognition of the role played by the body and environment. In his most recent book, *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (*NBC*, henceforth), Clark uses a wide-ranging survey of current and near-future technologies to argue that we humans are becoming (in some cases, already are) essentially cyborgs: ‘thinking and reasoning systems whose minds and selves are spread across biological brain and nonbiological circuitry’ (p. 3). He has essentially two goals with this claim: first, to further broaden his extended mind philosophy; and second, to argue that our capacity to extend our minds in these ways is quintessentially human.

As Clark has argued elsewhere (e.g., Clark & Chalmers, 1998), he is quite literal about the nature of his ‘extended mind’ claims. One of his two central theses in *NBC* is that a ‘mind’ is a tightly interconnected collection of informationally reliable components, where those components might occur in a biological brain, but might just as well be in the body, the environment, or distributed across multiple locations. What matters for Clark is not the physical location of the component, but rather its availability and reliability for planning, decision, and action. As a concrete example, Clark would presumably argue that the mind currently writing this review is made up of at least four distinct components: a slow, active biological processor (that ‘decides’ what to write); brain-driven muscle memory (translating the brain signals to finger strokes); a computer program (in which various thoughts about the review are worked out); and a physical copy of *NBC* (with numerous margin notes and pieces of paper recording various past thoughts). For Clark, excluding the book or computer program from this mind represents an unwarranted biological chauvinism: the biological/non-biological dimension simply does not map onto inclusion/exclusion in a mind. The notes in my copy of *NBC* are of course different from my biologically

encoded memories about its content, but these differences are not of the sort to warrant excluding the former and including the latter in my mind.

And just as—in part, because—our minds consist of collections of components whose parts can change, the ‘self’ is a similarly fluid concept. As Clark describes our minds and selves, they are ‘just tools all the way down’ (p. 136). This use of the term ‘tools’ to describe the mind/self is potentially misleading, as it might suggest that he is leaving the door open for the brain to assume a privileged role as the cognitive director. Elsewhere, however, Clark makes his minimalist view about the self quite explicit: ‘there is no informationally constituted *user* relative to whom all the rest is just *tools*’ (p. 192; italics in original). In fact, in subsequent commentary about *NBC*, Clark has moved beyond minimalism to suggest a relatively eliminativist view of the self: ‘the self is a hastily cobbled together coalition of biological and non-biological elements, whose membership shifts and alters over time and between contexts’ (Clark, 2004, p. 177).

The extended mind thesis of *NBC* has appeared in other of Clark’s writings; the more novel thesis is that this ability to incorporate non-biological functional components into a mind is exactly what distinguishes humans from other animals in the world. Clark is quite explicit about this claim in the Introduction: ‘what is special about human brains, and what best explains the distinctive features of human intelligence, is precisely their ability to enter into deep and complex relationships with nonbiological constructs, props, and aids’ (p. 5). Incorporation of external components is not something unusual or alien for us, but rather, Clark argues, something that comes completely naturally. *We qua* humans need not be concerned about becoming extended, embodied cyborgs, precisely because we already are. Our brains and bodies are ‘*tailor-made for multiple mergers and coalitions* [with other informationally reliable components]’ (p. 7; italics in original). In part, this claim is based on Clark’s argument that our current technological advances are not qualitatively different from many other, more common, cognitive prostheses:

Human-machine symbiosis, I believe, is simply what comes naturally. It lies on a direct continuum with clothes, cooking (“external, artificial digestion”), bricklaying, and writing. The capacity to creatively distribute labor between biology and the designed environment is the very signature of our species, and it implies no real loss of control on our part. For *who we are* is in large part a function of the webs of surrounding structure in which the conscious mind exercises at best a kind of gentle, indirect control. (p. 174; italics in original)

As this passage indicates, Clark suggests that this capacity to construct our selves from both biological and non-biological components can be given an evolutionary explanation. The argument is roughly that even a limited ability to shift cognitive loads onto non-biological components would have given our evolutionary ancestors a significant competitive advantage. Unfortunately, Clark gives relatively little in the way of substantive argument for this evolutionary explanation. He essentially provides a ‘how possibly’ story that suggests how this capacity could have evolved, but gives no real evidence that this explanation is actually correct. This gap should

not be taken as a deep criticism, however, since determining the evolutionary story is not the goal. Clark is here taking an initial step: arguing that humans are, in fact, uniquely capable of assembling multiple non-biological and biological components for various cognitive tasks, and have been doing so for generations. Explanation of the origins of that capacity would presumably be the subject of future work.

NBC is a deeply impressive, *tour de force* catalog of various inventions and devices that extend or change our cognitive capacities, and much of the book is naturally read as providing an extended ‘argument by example’ for Clark’s pair of theses. *NBC* is clearly intended for a broader audience than academic philosophers and psychologists. Clark is an entertaining writer, and the prose is quite readable and accessible, far more so than the vast majority of academic writing. *NBC* succeeds admirably at presenting a ‘popular’, accessible argument for our extended minds and unique capacity to assemble them.

That being said, success at this goal may leave many readers of this journal wishing for more. On the philosophical side, many important distinctions are relegated to endnotes. As just one example, Clark quickly passes over the problem of how the semantic content of neural activity is fixed because it ‘would require a long philosophical detour’ (p. 123). As a result, those who were previously unconvinced by Clark’s arguments (e.g., Adams & Aizawa, 2001; Butler, 1998) are unlikely to change their minds upon reading this book. Similarly, those seeking a philosophically technical consideration of the theoretical landscape around Clark’s position would be better off focusing on his other published works. On the psychological side, there is perhaps the most glaring lacuna in the book: an almost complete absence of any discussion of developmental psychology. Essentially all of Clark’s examples and discussions are drawn from experiments or devices used with adults. Given that Clark is arguing that our minds are constructed or assembled from a diverse range of components and tools, one might naturally have expected him to focus some substantial content on the ways in which this construction happens during crucial developmental periods. Of course, the ethical standards for developmental research are significantly higher than for adult research, and so, for example, there will be very few (if any) implant experiments on children. But one would have hoped for significant discussions of current developmental research, and demonstrations of how his examples are consistent with our latest, best theories. Such discussions are almost wholly absent from *NBC*.

More generally, *NBC*’s focus on the metaphysical character and implications of Clark’s theses means that discussions of other *implications*—both philosophical and societal—are much less satisfactory. On the philosophical side, consider just one example: what are the consequences for notions of moral responsibility, given this new, distributed, often non-biological, self? For example, suppose a doctor uses a PDA-based expert system as a diagnostic aid, to the point where the component meets Clark’s rough criteria for inclusion in her mind. Is the doctor now absolved of the moral responsibility of responding to a car accident when out of the office, simply because she does not have ‘part of her mind’ with her? Or has she assumed a new moral responsibility: namely, to ensure that she ‘has her mind with her’ wherever

she goes? I believe that Clark's extended view of mind and self is certainly compatible with moral responsibility. However, some discussion of the implications of his metaphysical theses on other philosophical problems would have made this book significantly philosophically meatier (though at the possible cost of significantly shrinking the audience).

Clark also devotes an entire chapter of *NBC* to the ways in which the incorporation of novel technologies into our cognition could impact our society, touching on such issues as inequality, privacy, alienation, information overload, and control. His discussion is technically sophisticated and quite well-nuanced, and Clark's optimism is evident throughout the discussion. In fact, sometimes his optimism threatens to fall into wishful thinking. For example, Clark suggests that our cars might retain information about our driving habits, and thereby do a better job of assisting our driving; gradually, our minds would be (partly) composed of the computer in our cars. Insurance companies would naturally seek that information, and if it were shared, then good drivers would presumably get better rates. Clark suggests that 'if you choose not to share such information, you will simply not get the discount and the policy will be priced the old-fashioned way' (p. 164). This is a deeply optimistic view; a more pessimistic (realistic?) individual might naturally conclude that insurance companies would simply interpret my failure to provide information as a signal that I am a bad driver, raise my rates, and thereby punish me for wanting to maintain some semblance of privacy.

Nevertheless, Clark recognizes his own basic optimism with regards to potential societal problems. He is not here trying to solve all these issues, but rather to ensure that the debates recognize his theses: incorporation of external, non-biological components into our minds is not something anomalous in human history; it is a crucial part of the human condition itself. Moreover, Clark would arguably suggest (correctly) that we are in a similar position to people trying to solve the potential societal problems brought about by the printing press: any proposals will almost certainly fail to predict accurately the actual impacts of these new technologies.

Despite these shortcomings, *NBC* is an excellent book for its goal, particularly since many of these more technical issues have been addressed elsewhere by both Clark and other authors. Clark presents a powerful case that we humans are not 'ancient biological minds in colorful young technological clothes. Instead, ours are chameleon minds, factory-primed to merge with what they find and with what they themselves create' (p. 141). By providing an account of *why* our minds are extended—it is what we are 'designed to do'—Clark provides an important and powerful motivation for his view of minds extended out from brains, into bodies and their environments.

References

- Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43–64.
- Butler, K. (1998). *Internal affairs: Making room for psychosemantic internalism*. Dordrecht, The Netherlands: Kluwer.

Clark, A. (2004). Author's response to 'We have always been... cyborgs'. *Metascience*, 13, 169–181.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 10–23.

DAVID DANKS

*Department of Philosophy
Carnegie Mellon University
135 Baker Hall
Pittsburgh, PA 15213 USA
Email: ddanks@cmu.edu*

Computationalism: New Directions

MATTHIAS SCHEUTZ (Ed.)

Cambridge, MA, MIT Press, 2002

viii + 209 pages, ISBN: 0262194783 (hbk); \$40.00

Matthias Scheutz—a computer scientist at the University of Notre Dame—has put together this collection in order to defend computationalism from recent attacks. To a first approximation, computationalism is the view that mental capacities can be explained in terms of computations, namely, processes that generate output symbols by following appropriate rules that apply to input symbols and internal states. Ever since computationalism was formulated in the 1940s, many innovative and successful research programs in psychology, neuroscience, and AI have been based on some version of computationalism. But from the start, computationalism has been controversial.

The most widely accepted criticism is that computing is insufficient for generating consciousness. Given that there is no consensus on what consciousness is or how to explain it, it is hard to make this objection precise. Nevertheless, many find it hard to believe that computing is all that's needed for consciousness. Scheutz is not worried about consciousness, so let's leave consciousness aside. At any rate, computations, if not the whole story, might remain an important part of the explanation of mental capacities.

A second influential criticism is that computing is insufficient for the intentionality (aboutness, contentfulness) of the mind. Again, there is little consensus on what intentionality amounts to or how to explain it, but most people believe intentionality requires something over and above computing. During the past three decades, this belief has been one motivation for theories of content and intentionality. Some philosophers engaged in this project believe that computationalism should be supplemented by a theory of content, so as to obtain a more complete theory of mind.

Others deny that mental states are intentional, or at least that intentionality should play an explanatory role in a scientific theory of mind. Recently, some deniers of

intentionality have criticized computationalism for postulating computations endowed with intentional content (e.g., van Gelder, 1995). This objection is queer. Given the consensus that computationalism cannot explain intentionality by itself, the rejection of intentionality should be seen as a *boost* to computationalism. For those who postulate computations with intentional properties do so because they want to explain the intentionality of the mind. If there is no intentionality to explain, computationalists may simply postulate computations that have no intentionality and dispense with the difficult problem of intentionality. Ironically, some deniers of the explanatory role of intentionality have presented their thesis precisely as such a boost to computationalism (Stich, 1983). Needless to say, most people—regardless of whether they are committed to computationalism—still maintain that mental states are intentional.

Contemporary anti-computationalists, such as van Gelder, have leveled two further criticisms. One is that supposedly, computationalism neglects the temporal dimension of mental capacities. Computations can be defined in terms of the steps it takes to complete them, without specifying how long each step takes. By contrast, organisms must satisfy the temporal constraints their environments impose on them or else die before they reproduce. Presumably, the explanations for mental capacities must refer to temporal properties that explain how organisms behave in real time. Because of this, the objection goes, computation is not suited to explaining mental capacities. But this objection is no more effective than the previous one. Just because computations can be defined without temporal constraints, it doesn't follow that they must be so defined. It is quite trivial to specify a temporal scale for a computation, for example by specifying how long it takes to complete a computational step. In fact, computationalists have always included a temporal dimension in their computational explanations of mental capacities (e.g., Newell, 1990).

Another recent critique is that computationalism ignores that organisms are coupled to their environment. This objection is even less incisive than the previous two, because computing mechanisms can be coupled to their environments neither more nor less than other systems. In fact, there is a sub-field of computer science entirely devoted to computing mechanisms coupled to their environment—what in the trade are called 'embedded systems'.

Scheutz seems to be aware that these recent critiques of computationalism are misconceived—the book makes plain that there is more to computation than computationalism's critics have dreamt of. Nevertheless, Scheutz also believes that in order to properly answer these critiques, the very notion of computation on which computationalism is founded needs to be replaced. The new notion of computation, Scheutz says sensibly, should respect standard results from computability theory. It should also take computing to be 'concrete, semantic, embodied, interactive, and intentional', and hence would offer 'a much better chance of serving as a possible foundation for a realistic theory of mind' (p. x). Scheutz says little on how a new notion of computation is to be found. And given that Scheutz seems to agree that recent critiques of computationalism are misconceived, he should also explain why it

is preferable to address the intentionality of the mental by searching for a new notion of computation rather than by formulating a theory of content.

Of the authors in the collection, the only one who is explicitly concerned with a new notion of computation is Brian Smith. For a number of years, Smith has maintained that our traditional notion of computation is inadequate and needs to be replaced (Smith, 1996). In this essay, Smith reaches an even more radical conclusion: there can be no theory of computation, because 'computation is not a subject matter' (p. 51). If this makes you wonder what computability theory is about, Smith tells you that it is 'a mathematical theory of the flow of causality' (p. 43). In order to understand both computation and cognition, Smith writes, computer scientists and cognitive scientists 'must develop a theory of everything' (p. 53). For more details, Smith refers to his seven volume opus magnum, *The Age of Significance*. Smith has referred to *The Age of Significance* as 'forthcoming' for at least nine years (cf. the bibliography to Smith (1996), when it was listed as having five volumes). Even so, there is no evidence that *The Age of Significance* is coming out any time soon. If a theory of everything is what Smith is up to, perhaps he should be forgiven for taking some extra time.

In contrast to Smith, Stevan Harnad and John Haugeland limit their subject to the problem of intentionality. Haugeland argues that computation 'presupposes' intentionality, for that is the only way computation could explain cognition (p. 160). Haugeland does not further explicate his notion of 'presupposition'. This is a pity, for there is relevant recent literature on whether computational states are individuated by their semantic properties. Some of the participants believe that computational states are individuated by their functional properties alone, without any reference to their intentional properties. Haugeland does not discuss any literature, recent or past. Instead, he speculates on how science works, and concludes that science, and more generally cognition, or rather 'authentic intentionality', requires the capacity to assume 'authentic responsibility'. He adds that 'cognitive science and artificial intelligence cannot succeed in their own essential aims unless and until they can understand and/or implement genuine freedom and the capacity to love' (p. 174). Haugeland warns the reader: his conclusion is 'perfectly serious' (ibid.).

Harnad's proposal is more down to earth. He defends an internalist version of functional role semantics, according to which content is determined by appropriate functional relations between symbols, sensory inputs, and motor outputs. Internalism faces well-known objections to the effect that genuine content is not (entirely) in the head, and functional role semantics faces its own range of difficulties. Rather than answering objections, Harnad 'confesses' that he hasn't 'really provided a theory of *meaning* here at all, but merely a theory of symbol grounding' (p. 155). By that, he means a theory of how a robot can 'categorize its sensorimotor projections' (ibid.); that is, a theory of how 'to connect the proximal sensorimotor projections of distal objects and events to either the instrumental responses or the arbitrary names that successfully sort them according to what is adaptive for the . . . system' (p. 147). Harnad's believes his project is still relevant, because he sees it as a first step towards a theory of intentionality.

One way that anti-computationalists have missed their target is by placing too much emphasis on features of abstract formalisms, such as Turing machines, at the expense of the richness and complexity of computers. Aaron Sloman attempts to redress the balance by listing important features of computers, some of which are not shared by Turing machines but might be relevant to theories of mind. These features are the use of internal addresses for memory registers, the coupling with the environment via transducers, the handling of interrupts, the checking of one's own processes for errors, and multitasking. Curiously, Sloman doesn't mention what is perhaps the most fundamental difference between computers and Turing machines: computers have distinct and functionally independent input devices, memory units, and output devices, whereas Turing machines have one component—the tape—which plays all those roles at once. (This difference is fundamental because it provides much of the functional basis for the differences mentioned by Sloman.)

In rejecting the relevance of Turing machines to computer science and AI, Sloman goes too far in two respects. On the one hand, he states that 'Turing machines have little or nothing to do with computers as they are normally used and thought of' (p. 88). This is hardly fair; Sloman's list of computer features includes many that are shared with Turing machines, and others could be added. More importantly, as Sloman recognizes, uncomputability results that were established using formalisms such as Turing machines provide an upper bound to what can be computed by computers. On the other hand, Sloman states that 'the development of AI did not depend even historically on the notion of a Turing machine' (*ibid.*). On the contrary, there is plenty of evidence that Turing machines played an important role in the origin of computationalism and AI. I will only mention two representative examples. First, Alan Turing's foundational investigations played a deep role in his influential argument that computers can be programmed to be intelligent (Piccinini, 2003). Second, when Warren McCulloch and Walter Pitts formulated modern computationalism for the first time, they were 'treating', in McCulloch's words, 'the brain as a Turing machine' (Piccinini, 2004).

One of the most valuable contributions in the collection is a reprint of a paper by Jack Copeland (2000). Copeland uncovers one source of the tendency, which is persistent in the literature surrounding computationalism, to conflate computationalism and mechanism. Mechanism—the idea that mental capacities can be explained in terms of mechanical processes—is *prima facie* broader than computationalism—the idea that mental capacities can be explained by computations. For on the face of it, computations appear to be only one kind of mechanical process among others. Yet, Copeland shows that many prominent authors—computationalists and anti-computationalists alike—have appealed to the Church–Turing thesis and universal computing mechanisms to conclude that every mechanical process is computational, and thus that computationalism is the only mechanistic approach to the mind. Copeland calls this line of argument the Church–Turing fallacy, and argues persuasively that it is based on a misunderstanding of the Church–Turing thesis. The Church–Turing thesis, Copeland reminds us, does not apply to mechanical processes in general, but only to computations that follow an effective procedure.

Copeland's positive proposal, however, is less persuasive than his critique of the Church–Turing fallacy. He does not explicitly distinguish between genuine hypercomputations, namely processes that reliably generate values of functions that are not computable by ordinary computers, and other non-computational processes (such as, say, radioactive decay, digestion, or photosynthesis). Copeland simply uses the term 'hypercomputation' for any process that is not an ordinary computation. As a result, his suggestion that mental processes might be explained by hypercomputations—albeit original and provocative—is less clear and less motivated than it could be. Nevertheless, his critique of the Church–Turing fallacy should be internalized by anyone interested in computationalism.

In sum, *Computationalism: New Directions* does not establish a new notion of computation, or even the need for such a notion. But the book contributes new and thought-provoking ideas to the ongoing debates that surround computationalism.

References

- Copeland, B. J. (2000). Narrow versus wide mechanism: Including a re-examination of Turing's views on the mind-machine issue. *The Journal of Philosophy*, 96, 5–32.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Piccinini, G. (2003). Alan Turing and the mathematical objection. *Minds and Machines*, 13, 23–48.
- Piccinini, G. (2004). The first computational theory of mind and brain: A close look at McCulloch and Pitts's 'Logical calculus of ideas immanent in nervous activity'. *Synthese*, 141(2), 175–215.
- Smith, B. C. (1996). *On the origin of objects*. Cambridge, MA: MIT Press.
- Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge, MA: MIT Press.
- van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92, 345–381.

GUALTIERO PICCININI
Department of Philosophy
University of Missouri – St. Louis
599 Lucas Hall (MC 73)
One University Blvd.
St. Louis, MO 63121-4499 USA
Email: piccininig@umsl.edu

Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition

WILLIAM D. CASEBEER
Cambridge, MA, MIT Press, 2003
224 pages, ISBN: 0262033100 (hbk), \$35.00

It isn't easy to naturalize ethics. It's even harder when one attempts to do so by drawing on such diverse disciplines as Aristotle's virtue ethics, John Dewey's pragmatism, evolutionary theory, connectionism, and neuroscience. Nonetheless, William D. Casebeer makes an ambitious effort to do just that in his recent

monograph, *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*. His account is a ‘fully naturalized’ one in which ethical norms are not only constrained but indicated by our evolved social natures. The volume is sweeping in scope and synthesizes recent advances in biology and cognitive science of potential relevance to naturalizing ethics; in spite of its sweep, however, the volume is relatively short. Casebeer often finds himself apologizing for his ‘brief’ treatments of the significant number of topics he attempts to cover. One main thesis is that facts about the roles that functions played in our modern evolutionary histories are also moral facts about the roles those functions should play in our contemporary lives. He also argues that our evolved capacities to make moral judgments and construct moral theories rest on neurophysiological processes that are best modeled using connectionism.

Casebeer begins by summarizing and attempting to respond to concerns over the naturalistic fallacy and the open question argument—concerns that most ethical theorists have whenever naturalism and ethics are uttered in the same breath. Such theorists worry that ‘ought’ cannot be derived from ‘is’; that moral facts cannot be derived from natural facts; and that whether a particular natural fact is also a moral fact is always an open question. Casebeer responds to these concerns by arguing that both G. E. Moore’s and David Hume’s arguments against the possibility of naturalizing ethics rest on an implicit analytic/synthetic distinction—a distinction clearly dispensed with by W. V. O. Quine. Casebeer writes, ‘With Quine in hand, we can insist that any *a priori* attempt to isolate the good from natural definition dodges tough questions about theory change: rather than insist that the meaning of good precludes natural definition, why not admit that you have a theory of the good (rather than merely a definition of it), and let such a theory be adjudged as theories are: by their relationship to other theories, and by their encounters with experience?’ (p. 26). According to Casebeer, Dewey similarly ‘eschews supernaturalism about the ethical and roots moral concerns in the activity of people coping with an environment’ (p. 27). Thus, Casebeer hopes to show that ethical norms and theories are pragmatic products of the experiences of evolved human beings, nothing more, and that there are better and worse theories depending on how well they contribute to our flourishing.

There is no doubt that his book would be much less controversial if he had stuck to ways that natural facts *constrain* our moral theories. However, Casebeer makes the much more controversial claim that moral facts *are* natural facts, or more specifically, that ‘moral facts are functional facts’ (p. 53). Roughly, his view is that an Aristotelian virtue ethics resting on a teleological view of humans can be updated using recent evolutionary biology and a modern-history theory of functions. Casebeer maintains a teleological view of humans by claiming that humans have proper biological functions and asserts that ‘by reducing moral terms to functional terms, and by treating the objects to which those terms refer as a contemporarily informed Aristotle would, I established a case for the objectivity of moral value . . .’ (p. 56).

Unfortunately, there is no consensus on either the concept or the role of functions in contemporary biology. Casebeer addresses this problem by attempting to place

his own views in the middle of what he refers to as a 'spectrum' of approaches to functions. The two endpoints of this spectrum are Larry Wright's etiological approach and Rob Cummins's capacity approach. He wishes to avoid the implications of a deeply historical approach to functions because on that view a human's function is simply to survive and reproduce and thus moral traits are simply those human characteristics that lead to successful reproduction. But he also wishes to avoid the implications of an ahistorical capacity approach because on that view 'anything goes' and moral traits are entirely relative. In hopes of getting a view of biological function that works for his purposes, Casebeer draws on what he terms a 'Wright-style analysis advocated by Peter Godfrey-Smith' that relates the function of traits to their recent evolutionary histories and supplements this analysis with Ruth Millikan's (1984) account of 'proper functions'. On Casebeer's view, such an approach to functions works best for updating Aristotle's ethics. It is somewhere between an approach in which functions are merely the current capacities of a system and an approach in which functions are so historically distal that they simply involve successfully replicating genes. As Casebeer puts it, 'Morally speaking it is not true that anything goes, but neither is it true that our only proper function in life is to breed like rabbits' (p. 53).

There are several reasons why Casebeer's account of functions and their role in determining norms is problematic. First, for biologists, functions play an *explanatory* role. Paul Davies, author of *Norms of Nature: Naturalism and the Nature of Functions*, puts it this way: 'The attribution of selected [historical] functions is equivalent to an explanation of why the functional trait persisted or proliferated and the attribution of systemic functions explains how certain systems exercise certain of their capacities' (2001, p. 31). The Godfrey-Smith account to which Casebeer appeals is a pluralist account rather than a single, middle-of-the-spectrum account. Godfrey-Smith (1994) actually argues that the systemic and historical approaches are each appropriate depending on one's explanatory aims. Sometimes it is one and sometimes the other. So it is unclear at best how Casebeer is drawing on Godfrey-Smith in straddling these two approaches. But even if this were clear, Casebeer doesn't have *explanatory* aims; he has normative aims. He wishes to say that because a functional trait has persisted over our modern history, it *should* persist. Surely there is a robust modern evolutionary history of aggression and its contribution to our evolutionary success. But why would that mean that the exhibition of this trait is necessary for human flourishing?

Casebeer does not devote much of the volume to describing which functions in particular are moral functions or to fleshing out his functional claims, but he does offer two very brief illustrations of how his functional approach provides us with specific normative content. I will focus here on the first of his illustrations that describes and solves a moral dilemma faced by 'John'. Because of his job, John often relocates. John's dilemma is whether it is more moral for him to 'cultivate many friendships that are all fairly shallow' or whether he should 'focus on cultivating a few deep friendships that might stand the test of time and the stress of relocation' (p. 141). Casebeer solves the dilemma by appealing to evidence that in our modern

evolutionary history, ‘humans were in intimate contact with a fairly small group of others (close relatives and kin, primarily)’. Thus, ‘some of the social capacity that we have probably has the proper function of enabling us to develop deep and intimate friendships’. According to Casebeer, once we identify this proper function of John’s social capacities, the moral dilemma is solved: John should focus on a few deep friendships. But I think this example only further illustrates the problem with Casebeer’s functional approach. First, isolating various aspects of human psychologies and in particular our social psychologies is a very different enterprise from isolating eyes and hearts. Normal humans have widely ranging types of social psychologies. Some are shy. Some are extroverts. Which is John and shouldn’t this factor into what types of friendships are best for John to pursue? Are there identifiable components of our social psychologies that we all have in common and that have a proper function? And even if we figure out their respective proper functions, why should these functions be normative? No doubt, in some sense, our ancestors had psychologies that allowed them to survive and even thrive in small groups. These same tribal ancestors also had psychologies that often led them to murder those who were not members of their small groups. Does this mean that our contemporary social interactions and structures should encourage tribalism and that one of our proper functions includes a psychological propensity to murder others? I actually think that Casebeer would have had a better chance at getting a functional account to work for him, if he had chosen to develop a capacity (ahistorical) approach to functions. He would need to start with a contemporary account of human flourishing and determine what range of psychologies (psychological functions) best contributes to this flourishing. But notice that identifying the norms for human flourishing would then come first and identifying the contributing functions would come second. Even here, I’m pessimistic. Biological functions just don’t work for ethics and for accounts of human flourishing.

Casebeer is well-aware that many readers will be unconvinced by his attempt to fully naturalize ethics using a functional approach, so he asserts early on that his project is nonetheless important because ‘we can at least maintain that the biological and cognitive sciences can constrain moral theorizing by identifying the realistic limits of our biological and moral capacities’ (p. 33). Casebeer devotes several chapters to the less contentious task of characterizing moral cognition. He draws on recent developments in cognitive science to describe how moral judgments are made and moral lives led. These sections of Casebeer’s book are more modest and more on track. Using a connectionist model and drawing on Paul Churchland’s recent work, Casebeer emphasizes that much of cognition including the cognition involved in the making of moral judgments is not linguistically mediated. Casebeer stresses the roles of mental models, analogy, and metaphor in our moral judgments and the role of developing moral perceptual and behavioral habits in ‘training up’ moral psychologies. I think Casebeer’s views here are basically right and that there are significant implications for these views in both how we theorize moral cognition and how we develop norms for moral cognizers. Casebeer is also right, I think, in his

controversial claim that non-human animals can make moral judgments of a sort. I hope he develops these significant ideas more fully in future work.

In summary, I recommend that Casebeer's volume be viewed as a good starting point for someone wishing to naturalize ethics using the latest research in cognitive science. It does have a very nice bibliography and provides us with an interesting combination of issues and ideas to be further explored.

References

- Davies, P. (2001). *Norms of nature: Naturalism and the nature of functions*. Cambridge, MA: The MIT Press.
- Godfrey-Smith, P. (1994). A modern history of functions. *Noûs*, 28, 344–362.
- Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, MA: The MIT Press.

PEGGY DESAUTELS

Department of Philosophy

University of Dayton

300 College Park

Dayton, OH 45469-1546 USA

Email: Peggy.DesAutels@notes.udayton.edu

The Lord of the Rings and Philosophy: One Book to Rule Them All

GREGORY BASSHAM & ERIC BRONSON (Eds)

Chicago, IL, Open Court Publishing Company, 2003

336 pages, ISBN: 0812695453 (pbk); \$17.95

Some readers of *Philosophical Psychology* will have been pleased when they heard that a book has recently appeared with the intriguing title, '*The Lord of the Rings*' and *Philosophy*. For *The Lord of the Rings* is a work which offers rich veins to mine, for philosophical psychologists. It is obviously a 'quest' book. It is natural to read it psychologically, then, as an account of a psychic journey as much as or rather than a real one. It offers psychological insights, when thus read, which sometimes attain the same depth as the psychological insights of other great 'quest' epics, such as the Arthurian legends, and Homer's epics. Like *The Odyssey*, for instance, it is a rich magical psychic trip centered above all on one man (or, in the case of *The Lord of the Rings*, on one hobbit, Frodo), and posing/dealing at crucial points with paradoxes involving the temptation of that man; like *The Iliad*, it can easily be read as an allegory of a psyche under trial—in the case of *The Iliad*, the trial is perhaps primarily one of decision over issues of honour and of practicality; in the case of *The Lord of the Rings*, it is perhaps primarily one of anxiety and the various routes to dealing with it. *The Iliad* has been interpreted as yielding a fascinating though dubious historical philosophy of psychopathology by Julian Jaynes; recently, it has been given a brilliant

more deflationary psychological reading by Leudar and Thomas. What of *The Lord of the Rings*?

The Lord of the Rings argues that the desire to achieve safety through the acquisition of power over one's fellows, one's life, one's experiences, leads only to self-defeating fantasies, and that the hard route of ordinary 'faith', and renunciation of any quick-fix fantasy of safety, is the only route that will succeed. The routes to dealing with extreme anxiety—anxiety that makes one desperate for safety—explored by *The Lord of the Rings* are: (1) a retreat away from the consensual world and deeper inside the mind (after the fashion of the patients central to Laing's *The Divided Self*), a journey to the edge of psychosis, symbolized by the putting on of the Ring (and also arguably by retreat within the armed cities of Helm's Deep and Gondor); (2) a 'giving in' to psychosis, a psychotic break, symbolized by the longed-for giving away of the Ring to the alienated part of oneself played by Sauron (the Lord of the Rings) and his surrogates; (3) a breaking of the power of one's anxieties over one, through ordinary-faith-in-action renouncing the fantasized power of one's would be anxiety-remedies, symbolized above all by the dissolving of the Ring back into the fires from whence it was forged; (4) the contemplation of (1) through (3), enacted for example by a thoughtful—one might say, a philosophical—reader/reading of *The Lord of the Rings*. Tolkien's book argues for (3) and (4), and against (1) and (2). In the course of doing so, it dramatizes and indeed investigates many philosophical issues of intense related interest.

For example, Sauron and his wraiths are the malicious demon of Descartes. They are the terrifying possibility that the world is godless, but for a god who is *malevolent*. They bring out what Descartes entirely fails to bring out: that the possibilities of philosophical skepticism are terrifying, an apotheosis of anxiety. The rapidity of Frodo's encounters with the lidless eye, the searching, penetrating, judging 'eye' of 'Sauron', is a testimony to the rapidity of the possible movement to psychosis in the mind of a quick-thinking 'meditator'. And psychosis can be simply lived solipsistic skepticism, as Louis Sass has argued.

How much of this is understood by the authors in Bassham and Bronson's book? Sadly, very little. There is some intelligent discussion of the non-existence of pure evil, i.e. the non-existence of Sauron, via Tolkien's Augustinian moral-philosophical worldview, for instance in Scott Davison's essay on 'Tolkien and the nature of evil'. Evil as lack is also well-explored around p. 174 of Thomas Hibbs's 'Providence and dramatic unity in *The Lord of the Rings*'. But in neither case are the psychological and psychopathological implications investigated. The crucial implicative connection between there not being true evil and the necessary human—moral and practical—task of not giving up on anything human, which lies at the root of the pity and fellow-feeling toward Gollum crucially demonstrated by Frodo (and by Bilbo in *The Hobbit*), is not pointed out. The purely spectral character of the enemies and monsters of *The Lord of the Rings* (which incidentally offers the only possible acquittal of the book from the charge of racism) is not followed up on. The Lord of the Rings is an apotheosis of evil. The Lord of the Rings does not exist. But what the Ring stands for—the fantasy of Lordship over reality—continues to exist, so long as

the human desire for it persists, so long as ordinary faith and life is 'not enough'. This is in part what is symbolized by Frodo's inability to live in the earthy earthly paradise of The Shire at the end of *The Return of the King*: the extraordinary difficulty of recovering from the wounds sustained in the course of a psychotic journey. Frodo's deepest wound never fully heals.

There is some bright discussion of Gollum's 'split personality' (e.g. around p. 66 of Jorge Gracia's 'The quests of Sam and Gollum for the happy life'); but no investigation of Gollum as our putative nightmare, a living moral-nihilist, as our phantasized reduction of life to mere biological life, craving, pure addiction. Gollum lives in perpetual paranoid fear of Sauron; he is the image (though not the reality, for he is not lost—no-one is) of what we (Frodo) fears becoming; if he can get lost in madness, then so can we, for he was like us, once. Again, even this relatively obvious psychic and moral dimension of the text is largely absent from this collection.

There is some intelligent discussion of Tom Bombadil and of the different sense of time that is required by the ecological movement (in Andrew Light's 'Tolkien's Green Time'); but Light neglects to connect this with the ecological consciousness of the Ents, intermingled with that of Merry and Pippin, expressed in their unwillingness, again for both moral and prudential reasons, to give up on any part of the natural Earth. An integrated mind is allegorized by Tolkien in the wholistic sense of the Earth and of its peoples that is writ large in the mutual aid that undergirds the unlikely victories—unlikely in the sense of being achieved in the face of the *overwhelming* might of the enemy (the overwhelming urge to escape or retreat psychically, to give in to a flight of anxiety, being again psychologically central to Tolkien's text, and still more so to Jackson's recent film trilogy)—in the spectacles of dread and its overcoming at Weathertop, at Helm's Deep, at Orthanc, at Gondor and on the Pelennor Fields, and so on.

Most crucially, there is no discussion at all of the points made earlier about what Sauron does to enrich our understanding of what Descartes's demon is, about what the Ring means in psychopathological terms, nor about the advantages and disadvantages contemplatively and therapeutically of experiencing *The Lord of the Rings*.

The title under review is then a lost opportunity, and is moreover in parts almost as risible as its subtitle, *One book to rule them all*. It cannot be recommended, except as light reading, or as reading for philosophers-lite.

Which is a great shame. *The Lord of the Rings* still awaits a proper philosophical treatment. Its often subtle and powerful philosophical psychology deserves to be perspicuously presented to a public that I think senses it better than do a number of the authors in this collection.

RUPERT READ

Department of Philosophy
University of East Anglia
Norwich, NR4 7TJ, UK
Email: R.Read@uea.ac.uk