Did My Neurons Make Me Do It? Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will. By Nancey Murphy and Warren S. Brown. Oxford: Oxford Univ. Press, 2007. xviii +334 pages. $75.00.

According to some philosophers of mind, the pernicious influence of Cartesian dualism between mind and body survives among many contemporary materialists. These "Cartesian materialists," according to Daniel Dennett's terminology, claim to banish Cartesian minds from existence, but by identifying minds with brains they succeed only in embracing a new dualism of brain and body. This holdover from Cartesian dualism, some argue, has led to, among other vices, an unwarranted modularization of mental faculties, a myopic exclusion of factors external to the brain, and an unhealthy moral solipsism.

To begin the return to philosophical virtue, Nancey Murphy and Warren Brown suggest that we must reject the dominant intuition that mental states are somehow located "inside" our brains. Instead we ought to think of mental states as socially embedded brain states rooted in a larger physical world. Their book is devoted to the development of this thesis. The problems with Cartesian materialism are laid out in the first chapter. An account of mental states construed as contextualized brain states is covered in the next four. The final two chapters work out the consequences of this view for moral responsibility.

According to Murphy and Brown, the crucial step in eradicating Cartesian materialism is to confront causal reductionism—the thesis that the behavior of an entity is fully determined by the behavior of its parts. The picture of the world that emerges from causal reductionism is a hierarchical organization of systems based on levels of complexity where lower-level entities are the constituents that form the entities found at the next higher level. Being nothing more than aggregations of lower-level entities, however, the causal powers of higher-level entities are reducible to the causal powers of their components. By tracing the hierarchy down to the lowest, fundamental level, causation ultimately is located in the interaction of atoms. Consequently all causation is bottom-up, and only the isolated forms of causation taking place among the atomic components of the brain are needed to explain the mind.

Given this characterization, it is no surprise that Murphy and Brown believe that the demise of causal reductionism depends on the articulation and defense of top-down causation. According to them, causal accounts that are limited to lower-level laws are inherently incomplete, and developing an account of top-down causation will be a matter of fixing the "other factors." As a result, top-down causation is not something "spooky" that violates the lower-level laws; rather, higher-level wholes causally *constrain* the behavior of their lower-level parts (pp. 64–67).

Murphy and Brown cite evidence from biology, information theory, and cybernetics, among other disciplines, to illustrate how a larger system can constrain the behavior of its component processes. From biological entities that can be explained only functionally in relation to the larger systems in which they are embedded, to self-sustaining organizational patterns that are decoupled from their constituents, to nonlinear systems that exhibit holistic phenomena that are underdetermined by a mere summation of their parts, they claim that instances of

top-down causation are legion. Using Donald MacKay's action-feedback-evaluation-action (AFEA) model, they believe a plausible account of top-down causation can explain how a system pursuing its own self-maintenance "exerts constraints on [its] own components [and is] capable of selecting the stimuli in the environment to which [it] will respond, making [it] semi-autonomous from environmental control . . . and becoming (in part) [its] own cause" (p. 90).

To get clear on how this all works, let us take a closer look at one of their examples: a thermostatically controlled heating system. The system sustains the room's thermodynamic equilibrium by constraining its component processes. The furnace, S, has a causal repertoire of two states: on or off. Its current state, however, cannot be explained without taking into account the constraints imposed by a larger system, S', that includes the thermostat. The thermostat in turn has a causal repertoire of forty-one states: 50 to 90 degrees. The thermostat's current state, however, cannot be explained without reference to an even larger system, S'', that includes the entire room. One lesson to be learned from this example is that context plays an indispensable role in the causal activities of the components. Another lesson is that S'' can be characterized as self-maintaining. Its parts have degrees of freedom that are constrained by the system as a whole and, in maintaining a certain temperature despite a varying environment, it achieves autonomy. It acts, receives feedback, evaluates the feedback, and adjusts in an unending cycle.

If this model is correct, say Murphy and Brown, it helps dispel an intuition that contributes to the recalcitrance of the problems surrounding mental causation. It shows that organisms are not fundamentally passive entities that are waiting on external stimuli or an "inner" ego to exert its will. Organisms are continually active, and behavior is emitted from the organism as a whole. The causal role of the mental is one not of initiation but of ongoing behavioral modulation. Hence, the mind is what occurs when the brain, operating as part of an active system, engages the world.

The AFEA model is central to the authors' overall project. With modifications and extensions, it serves as the foundation for analyses of the troublesome features of mentality. To understand the possibility of intelligent action, Murphy and Brown begin with a bare-bones AFEA model to analyze lower organisms. Single-celled organisms have the ability to act reflexively. Protozoa, for example, that detect chemical concentrations in liquids and evaluate their toxicity can then act by moving away from deleterious regions. By adding a bit of short-term memory and a rudimentary supervisory system, more complex organisms are able to modulate activity, allowing them to achieve "*unreflective* adaptive action." When existing AFEA loops are nested within other loops that make meta-evaluations possible, organisms gain the ability to use representations and consequently achieve "*reflective* adaptive action" (p. 120). In higher organisms, such as humans, representations can be turned inward, making it possible to execute off-line simulations of potential behavior that are essential for intelligent action.

Representation itself is analyzed through a causal theory of reference. A given perception, say, of a carrot, actually refers to a carrot, because the perception is a "conditional readiness to reckon" with the perceived object (pp. 152–54). This implies that representations are stored not in isolated brains but in entire brain systems whose function is to initiate action. It is the deployment in action that is crucial to the authors' understanding of representation. They supplement this

analysis with evolutionary biology in order to account for the possibility of misrepresentation. By selecting the functions of representations that improve survival rate, evolution effectively imputes a normative role on the way representations ought to be used. In fixing the normative function of a representation, room is made for representations to go awry and consequently misrepresent. Applying these general insights to the issue of language, they give an account of how utterances take on meaning that respects the Wittgensteinian requirement for public accessibility.

Finally, the authors tackle the problem of rationality—that is, how mental events qua reasons have causal effects in the physical world. Although this problem may have seemed intractable due to the irreducibility of reasons with respect to neurobiology, the solution to the problem becomes straightforward when analogized to the thermostatically controlled heating system. The metal coil in the thermostat is sensitive to fluctuations in temperature, but the different states of the coil cannot count as *information* unless it is embedded in a system that can use that information to activate or deactivate the furnace. In other words, information supervenes on the coil embedded in the thermostatic system and not on the coil alone. In the same way, reasons cannot reside in isolated brain states and are thus irreducible with respect to them. They reside in total brain systems that are poised for action. By taking the entire organism into consideration reasons figure naturally into the network of causal relations.

With an account of top-down causation in hand, Murphy and Brown move to a discussion of moral responsibility. Their analysis relies in large measure on a broadly MacIntyrean understanding of moral responsibility based on voluntary action grounded in a capacity to evaluate one's own reasons. On this view, morally responsible behavior requires, among other things, robust representations of the world and of the self. These representations, which are accounted for in terms of nested AFEA loops, allow for the off-line simulation of potential behavior. These simulations provide an arena in which evaluative assessments can be made of various reason/action pairs. Furthermore, these assessments are not carried out in isolation but are informed by the rich ontogenetic landscape of the organism's history of actions. What this provides, by the authors' lights, is a plausible account of human voluntary action.

Although Murphy and Brown are satisfied with this account of moral responsibility, they anticipate a rejoinder from opponents who claim that the most important feature of moral responsibility has been left out: free will. To address this they suggest that the debate itself is improperly framed in terms of an illusory tension produced by deterministic and indeterministic accounts of freedom. Instead of looking at when and how particular actions are wholly due to an agent, Murphy and Brown focus on the claim that agents are *constantly* engaged in action. Self formation and re-formation is a never-ending process. Instead of searching for the "ultimate" source of responsibility and deciding whether this source is compatible with determinism, it is enough to locate the "primary" source of responsibility found in the AFEA loops that make agency possible. In this way they try to diffuse the free-will rejoinder.

This concludes a brief summary of their book. I now turn to a few reservations. While Murphy and Brown acknowledge the importance of consciousness and are keen on pointing out the essential contribution it makes to certain ac-

tions, I am suspicious that they have somehow dodged the primary issue. They seem to think that consciousness is exhausted by the causal role it plays in providing information relevant to action. They cite Weiskrantz's famous "blind sight" study to point out that among the things missing in people with blind sight is a kind of second-order knowledge—that is, knowledge of their visual knowledge. Characterized as second-order knowledge, consciousness is easily integrated into their AFEA model by adding further feedback-evaluation loops. They claim that the presumable rejoinder—What about someone who has second-order knowledge but no conscious awareness?—is an "incoherent" question (p. 220). But why is this incoherent? Surely a computer system can be designed such that it instantiates a relevantly deep hierarchy of AFEA loops that includes "monitoring" meta-loops without our thinking that the computer was thereby conscious. On their view, sufficiently sophisticated computers would also enjoy consciousness, and it seems an unpalatable conclusion that consciousness comes so cheap.

Part of the problem is that they conflate consciousness with self-consciousness. By thinking that "a person with sight knows that she is seeing something and also knows that she knows because she is conscious of seeing it" (p. 220) and analyzing consciousness in terms of self-consciousness (or second-order knowledge), the mystery of consciousness itself is left untouched. It seems that the authors are committed to a "Higher Order Thought" theory of consciousness, which has been developed at length elsewhere (Rosenthal 2005), but they do not extend the debate in any interesting way.

Consciousness aside, top-down causation, which is central to Murphy and Brown's program, is my greatest concern. They point out that top-down causation is really a constraint-based causation that is context-sensitive. Accordingly, they are critical of simplistic versions of supervenience employed by reductionists that deal only with covariation between supervenient and base properties. Going back to the thermostat example, the metal coil used to gauge the relative warmth of the environment carries temperature information when it is properly embedded in a thermostat. Detached from the thermostat, the coil fails to carry this information. That is, temperature information fails to supervene on the coil absent certain contextual factors. According to Murphy and Brown, "property S supervenes on base property B if and only if x's instantiating S is in virtue of x's instantiating B under *circumstance c*" (p. 206; emphasis added). By neglecting context, reductionists cannot account for the fact that it is only when the coil is embedded in the thermostat that certain supervenient information properties are instantiated and become causally relevant.

This, however, seems to be an unfair assessment. Jaegwon Kim (2005), one of the reductionist philosophers the authors are keen on criticizing, is aware of the shortcomings of context-insensitive accounts of reduction, and his version of reduction seems to accommodate their concerns. He is, after all, careful in espousing species-specific reductions and is wary of making simplistic reductions across the board. His is a *functional* reduction that assigns a causal role to mental properties. Being a role, it is necessarily context-involving—that is, roles must properly be embedded in a larger system that make the input/output specification possible. Once this role is defined and the necessary causal mechanisms that realize this role are identified, the reduction is made. In the end, what is doing the causal work, according to Kim, is the mechanism. That is, the functional prop-

erty that the mechanism realizes just *is* the mechanistic property that carries out the role. This account is indeed sensitive to context; take the mechanism out of its proper environment and the functional properties disappear. So it seems Kim's account is able to account for the supposed context-sensitivity of top-down causation. However, insofar as functional reductions alleviate worries over the causal exclusion of supervenient properties, his account enjoys the benefit of eschewing lingering epiphenomenalist worries.

More pressing, their account of top-down causation borders on vacuity. The authors stress that top-down causation does not violate any lower-level laws. But to maintain a nonvacuous account of top-down causation they must insist that there are gaps in the domain of lower-level entities that are left unaddressed by the lower-level laws alone. Indeed, they claim that the lower-level laws cannot tell the whole causal story—that bottom-up causation is inherently incomplete. When one brain state causes another brain state, the behavior of their lower-level constituents cannot be fully explained by their obedience to lower-level laws. But what precisely is left out of the bottom-up picture? Is it the "*organization* of [the] constituents within [a] composite" (p. 66)? Is it the massive interconnectivity among the components of dynamic systems (p. 72)? Is it the development and modification of behavioral ontogenetic landscapes (p. 77)? Is it sensitivity to "higher-level instructions" (p. 98)? I have difficulty in seeing how all of these higher-level features cannot be placed under the reductionist regime. Their examples of top-down causation always cite context as crucial in constraining lower-level behavior, but I cannot see how the context itself can resist a reductive gloss in lower-level terms.

Even if Murphy and Brown are correct in claiming that the bottom-up picture is incomplete, it is unclear that AFEA loops can capture the distinctive features of top-down causation. Why can't the behavior of the "comparator," "organizer," and "supervisor" (pp. 128–31), among other subsystems crucial to AFEA loops, be reduced to the behavior of their lower-level parts? A case in point is the modern-day computer—a paradigm example of the AFEA architecture. In fact AFEA loops mirror (and are even isomorphic to) schematics used by computer scientists and engineers in developing the various components of a computer system. When one stops to think about how bits of silicon trading in the currency of determinate voltage drops can provide an environment for a person to read an electronic paper, manage a database, or play a virtual 3-D game, it is easy to get seduced by the astonishing higher-level features. But does anything in a computer resist reductive analysis? Indeed, it is the very insights gained through the manipulation of lower-level entities based on exhaustive knowledge of lower-level laws that provide the theoretical underpinnings for developing computing systems.

While there is no doubt that computers often are given a high-level interpretation in terms of systems (processor, memory, cache, disk, I/O device etc.), this is only because a more fine-grained analysis often is cumbersome and unhelpful. If this, however, is all the authors mean by adopting a systems approach, it does not pose a threat to reductionism because it is not, in the end, a metaphysical proposal. In fact, Murphy and Brown seem to suggest that their approach is an epistemological endeavor when they write that adopting top-down causation is "something akin to a . . . Gestalt switch" (p. 43)—a shift in perspective to facilitate our explanatory needs. By shifting from an analysis of causal *efficacy* (chapter

2) to an analysis of causal *relevance* (chapter 5) they seem to espouse an account akin to Jackson and Pettit's program explanations (1990) that was intended to save higher-level causal *explanations* while denying their efficacy. The worries raised by reductionist philosophers such as Kim concerning the causal efficacy of mind are not epistemological in nature; the real issue for them is the seeming paradox generated by the causality of an irreducible mind embedded in a materialist metaphysics.

Murphy and Brown have tackled a gargantuan topic and have attempted to address the major issues facing theorists of mind and action. The strength of their undertaking lies in their ability to integrate large bodies of disparate knowledge. Putting the deliverances of neurobiology, psychology, information theory, and philosophy into a coherent system is no easy task. What they may have missed in terms of deep and probing analyses they surely make up in terms of their expansive vision. Although I would have liked more engagement with the extant theories of consciousness and have worries concerning their account of top-down causation, the book is clear, well-written, and valuable for its extensive interdisciplinary work.

References

Jackson, Frank, and Philip Pettit. 1990. "Program Explanation: A General Perspective." *Analysis* 50:107–17.

Kim, Jaegwon. 2005. *Physicalism or Something Near Enough*. Princeton: Princeton Univ. Press.

Rosenthal, David. 2005. *Consciousness and Mind*. Oxford: Oxford Univ. Press.

Daniel Lim
12327 Essex Street
Cerritos, CA 90703

Cosmic Jackpot: Why Our Universe Is Just Right for Life. By Paul Davies. London: Penguin, 2006, and Boston: Houghton Mifflin, 2007. xv + 315 pages. $26.00.

Paul Davies is a celebrated cosmologist with a sustained interest in its philosophical dimensions. Years ago he wrote *God and the New Physics* (New York: Simon and Schuster, 1983), and in the quarter century since he has repeatedly returned to themes surrounding the anthropic principle. *Cosmic Jackpot* is the sequel, maybe even a finale. If he can answer the question in his subtitle, that will really hit the jackpot.

Cosmology is changing (dark energy, dark matter, the thermal birth map of the universe, made with the Wilkinson Microwave Anisotropy Probe), so the issue needs revisiting. Davies is unsurpassed in summarizing highly technical results and their significance for a reasonably literate audience, such as *Zygon* readers. He can couple this with a conversational style, often with reminiscences about the cosmological celebrities involved.

He can be refreshingly blunt separating science from speculation and worrying about the transition zones: "As we consider earlier and earlier moments, we have to rely on increasingly speculative theories. Inflation, for example, makes use