Did My Neurons Make Me Do It?

By Nancey Murphy
Australia, August, 2011

[NOTE: This is the text of Nancey's lecture. It is not a formal paper and its referencing is incomplete.]

I had the opportunity ten years ago to visit Australia, and presented two papers titled "Why Christians Should Be Physicalists" and "How Physicalists Can Avoid Being Reductionists." This talk will focus on what I've learned on the second of these topics in the meantime. But first, I need to say something about physicalism (that is, the rejection of dualism of body and mind or soul) and why it is an acceptable position for Christians.

Then, in the main part of my paper I'll give a brief overview of where my thinking was on neurobiological reductionism ten years ago, but focus on dramatic changes since then. My interest in this topic is both philosophical and theological. Most philosophers of mind now count themselves as physicalists, and the main topic of discussion is whether physicalism necessarily implies reductionism. This is the thesis that all human thought and behavior are determined by the laws of neurobiology. In the past few years, many have judged that it does. Some philosophers, such as Daniel Dennett, are not bothered by this conclusion. He believes that we don't need the traditional concept of free will—we should think of ourselves as a system of trillions of mindless robot teams and nothing else. He should be bothered, though. If human thought is determined by neurobiology, then in what sense can it be determined by reason? Presumably he wants to provide rational persuasion for his point of view, but aren't we, on his account, simply determined either to believe him or not?

For Christian physicalists there are additional issues at stake: Can we still speak of our moral responsibility before God, or even of coming to believe in God, in part at least, on the basis of reason? Or do we have a new scientific version of predestination?

In the process of working on this issue, I've stumbled upon the resources for a significant shift in worldview. This shift has various and sometimes obscure points of origin, but like tributaries to a stream, these origins have now resulted in a flood of new ideas that are reshaping our understanding of causation. This new point of view, a systems-theoretic approach, is making us aware of the inadequacies of the modern mechanistic worldview derived from modern physics.

I will do my best to convey this new way of thinking, and then suggest, briefly, how it offers a solution to the problem of neurobiological determinism.

Sec. 1: Why Christians Should Be Physicalists

I claimed in my paper ten years ago that the Bible has no clear teaching on the metaphysical composition of humans—that is, are we bodies with spiritual capacities, or bodies and souls, or bodies, souls, and spirits. The lack of any specific answer to this question, I argued, leaves Christians free to adopt physicalism, which fits much better with current work in philosophy, and especially in neuroscience. Neuroscience is now locating brain regions and systems responsible for all of the functions once attributed to the soul or mind.

I would make a stronger claim at this point. Biblical views of human nature come closer to physicalism than to the dualist views of contemporary Christians. However, a potential rebuttal to my claim is this: If I and my fellow Christian physicalists are

correct, then how can we explain the fact that most Christians, throughout most of Christian history, took dualism to be the clear teaching of the Bible? I'll give an answer, but at two levels of generality. This more general response is that scholars have come to recognize that concepts have a history, and are also culturally specific. Therefore, to determine what the Hebrew word meant at the tie we can't assume that it means the same as our contemporary English translation does. For example, the Hebrew word *nephesh* was first translated into Greek as *psyche* and then into English as 'soul,' but we can't assume that it meant what we mean by 'soul' today, or what Plato meant by *psyche* in 300 BC.

This sort of rejection of essentialism with regard to language has led biblical scholars, throughout the whole past century, to engage in the detective work of situating anthropological terms such as 'soul,' 'mind,' 'heart,' and even 'body' in the thought-worlds of the biblical authors. Hebrew seems to work much more on the basis of the metaphorical extension of meanings than English does. Jewish scholar Neil Gillman says that the word *nephesh*, meant first of all the neck or throat, then by extension, the breath that flows through the throat. It was also used to refer to the life-blood. So by further extension *nephesh* referred to a living human being since it referred to the two characteristics that make a person alive, breath and blood (Gillman, 76).

A more specific way of understanding the difference between biblical views on human nature and our own comes from New Testament scholar James Dunn. He makes a helpful distinction between two different ways of using anthropological terms. In very general terms, classic Greek thinkers were interested in what he calls a partitive account of humans—that is, what are the essential parts that make up a human being. In contrast,

Hebraic thinkers thought of people primarily in terms of their relationships, and used anthropological terms aspectively—that is, to refer to the various aspects or dimensions of human life. Thus, even Paul's use of *soma,* body, does not function as we use 'body' today. Paul's concept has a range of meanings and includes the aspects of corporeality that enable our social relationships. Similarly, 'flesh' and 'spirit' do not refer to our later concepts of the material and immaterial parts of humans, but rather to two ways of living, one in tune with the Spirit of God, and the other opposed to such a way of life.

So I have a suggestion for getting a little bit closer to the minds of the biblical authors. Try replacing nouns such as 'mind' with characteristics such as rationality, 'soul' with living vitality, and so forth.

Sec. 2: The Problem of Neural Determinism

I've already stated the problem that the physicalist needs to solve: If human thought and action are the result of brain processes, how can it *fail* to be the case that we are all determined by the laws of neurobiology? And if we can't answer this, how can we account for moral responsibility, spirituality, or even rationality?

Now, I'm assuming that there are dualists in the audience, and so before you start playing games on your iPhones, I want to note that dualists have to reckon with this problem too. If what your brain does is determined by your neurons, how can your mind or soul make it do anything different? This is the problem, beginning with Descartes' work in the seventeenth century, of mind-body interaction.

Before proceeding, I want to frame the problem I'm addressing a bit differently. Current discussions of free will focus on neurobiological *determinism.* Two worries about

human determinism, genetic and neurobiological, are both instances of a broad thesis that may be called *causal reductionism*. Causal reductionism is the general claim that the behavior of all complex entities is determined by the behavior of their parts. This *is* the case in many of the systems we understand, such as mechanical clocks. These are designed so the movement of the parts determines the behavior of the whole. The problem is that there has been a tendency throughout the modern period to assume that when we turn to entities that are too complex to understand in detail, they also must be determined by their parts. In the human sphere, the significant parts are brain components, and so there has arisen a very sensible worry, namely that the laws of neurobiology are inevitably determining all of our thoughts and actions. So my focus will be not on neurobiological determinism, but rather neurobiological reductionism.

My colleague, neuropsychologist Warren Brown, and I worked for about seven years producing the book which is the basis for the title of this lecture: *Did My Neurons Make Me Do It? Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will.* When we started our work we believed that the answer to the problem of neurobiological reductionism was to develop and apply a concept of downward causation or whole-part constraint. That is, if causal reductionism in general is the thesis that all causation is from part to whole, then the complementary alternative would be whole-part causation. Alternatively, if we describe a more complex system, such as an organism, as a higher-level system than its biological parts, then causal reductionism is bottom-up causation, and the alternative is top-down or downward causation.

However, after pulling together all of the intellectual resources we could find to support the idea of downward causation, we read a book my Alicia Juarrero, titled *Dynamics in Action: Intentional Behavior as a Complex System.* It seems to me that about once every ten years or so I read a book that really changes how I think. This book has been one of the most important. First, we learned from Juarrero that even if you use downward causation to argue against neural determinism, this still doesn't give an account of human agency. That is, how are we to explain why we're not just passive players influenced from "above" by our environments and from below by our biology?

Second, Juarrero introduced us to complex-systems theory. In what follows, I'll first give a brief overview of work on downward causation, since it does play a very significant role in complex systems, and then I'll put it in the context of systems theory.

Sec. 3: Downward Causation

The earliest use of the term 'downward causation,' I believe was by Donald Campbell in 1974. Campbell provides an account of a larger system of causal factors having a *selective* effect on lower-level entities and processes. His example is the role of natural selection in producing the remarkable jaw structures of ants and termites. Bottom-up causation explains the mutations that alter the insects' genomes, as well as the building of the jaw structures. But in addition, he says, "[b]iological evolution in its meandering exploration of segments of the universe encounters laws, operating as selective systems, which are not described by the laws of physics and inorganic chemistry." "Where natural selection operates through life and death at a higher level of organisation, the laws of the higher-level selective system determine in part the distribution of lower-level events and

substances. Description of an intermediate-level phenomenon is not completed by describing its possibility and implementation in lower-level terms. Its presence, prevalence or distribution (all needed for a complete explanation of biological phenomena) will often require reference to laws at a higher level of organisation as well."

The best recent work I've found on downward causation is by Robert Van Gulick. He made an important contribution by spelling out in more detail than Campbell's an account based on *selection* (1995). The reductionist's thesis is that the causal roles associated with the classifications employed by higher-level sciences are entirely derivative from the causal roles of their underlying physical constituents. Van Gulick counters that even though the events and objects picked out by higher-level sciences *are* composites of physical constituents, the causal powers of such an object are not determined solely by the physical properties of its constituents and the laws of physics. They are also determined by the *organization* of those constituents within the composite; it is just such patterns of organization that are picked out by the predicates of the higher-level sciences.

These patterns have downward causal efficacy in that they can affect which causal powers of their constituents are activated. "A given physical constituent may have many causal powers, but only some subsets of them will be active in a given situation. The larger context (i.e. the pattern) of which it is a part may affect which of its causal powers get activated. . . . Thus the whole is not any simple function of its parts, since the whole at least partially determines what contributions are made by its parts" (1995, 251).

Such patterns or entities are stable features of the world, often despite variations or exchanges in their underlying physical constituents. Many patterns are self-sustaining

or self-reproducing in the face of perturbing physical forces that might otherwise destroy them (e.g. DNA patterns). That is, selective activation of the causal capacities of the pattern's parts may contribute to the maintenance and preservation of the pattern itself. These points illustrate that higher-order patterns can have a degree of independence from their underlying physical realizations and can exert downward causal influences without altering the underlying laws of physics (1995, 252).

Van Gulick's position, however, is open to the following objection: The reductionist will ask *how* the larger system affects the behavior of its constituents? To affect it must be to cause it to do something different than it would have done otherwise. Either this is causation by the usual physical means, or it is something "spooky." If it is by usual physical means, then those interactions must be governed by ordinary physical laws, and so all causation is bottom-up after all. This is the point at which we need to shift to the perspective of complex systems theory.

Sec. 4: Complex Systems Theory

The late Alwyn Scott was a mathematician, who applied non-linear mathematics to the understanding of neural processes. He claimed that the development of complex systems theory represents a paradigm shift across all of the sciences. Francis Heylighen goes even further in arguing that systems thinking will provide the basis for an entirely new worldview.

"Systems" thinking has been developing over the past half-century, although it has only recently begun to have a significant impact. Systems theory draws from a number of sources. As the term implies, there are significant roots in general systems

theory, developed from the 1950s through the 1970s by thinkers such as Ludwig von Bertalanffy. Bertalanfy was interested in explaining why it was the case that equations of the same form turned out to be applicable in widely different areas of science. Another early source was the study of cybernetics, which began as the study of feedback control processes in mechanical systems in the 1940s, and has turned out to be essential for understanding regulatory and goal-directed processes in biology. Current contributions come from information theory, nonlinear mathematics, the study of chaotic and self-organizing systems, and non-equilibrium thermodynamics. Examples of the systems of interest range from autocatalytic processes, at the most basic, to weather patterns, insect colonies, social organizations, and, of course, human brains.

I attempt to set out here some of the essential concepts involved in this change, and then illustrate them with the example of an ant colony. I will then report ever so briefly on work done with Warren Brown in tying systems theory to the assorted abilities that go into our capacity for morally responsible action.

Several authors call for what might be called a shift in ontological emphases. Juarrero says that one has to give up the traditional Western philosophical bias in favor of *things*, with their intrinsic properties, for an appreciation of processes and relations (1999, 124). Heylighen argues that the basic ontological categories for systems theory are agents and actions (2000, ts. pp. 6-8).

Systems have permeable boundaries, allowing for the transport of materials, energy, and information. The boundary is a matter of the tighter coupling of its components with one another relative to their coupling with entities outside of the system. The crucial components of complex systems are not so much things but

processes. So, for example, from a systems perspective, a mammal is composed of a circulatory system, a reproductive system, and so forth, *not* of carbon, hydrogen, calcium. The organismic level of description is largely *decoupled* from the atomic level.

Systems are different from both mechanisms and aggregates in that the properties of the components themselves are dependent on their being parts of the system in question. Philosophers distinguish between internal and external relations. External relations do not affect the nature of the *relata,* but internal relations are partially constitutive of the characteristics of *relata.* An essential assumption of the predominant modern worldview was that the world is composed of *things* related to one another *externally.* Systems theory takes the relations among the constituent *processes* of a system to be *internal.*

Systems range from great stability to wild fluctuation. This is due to the fact that complex systems are nonlinear, that is, the current state affects the development of each future state. The difference in stability is due to the extent to which the system is sensitive to slight variations in initial conditions, and also to the extent to which there are feedback processes that either do or do not dampen out fluctuations. Systems at the extremes of this spectrum of stability are not of great interest to systems theory. Consider as an analogy a thermostatically controlled heating system. It's very stable but produces no novelty because it involves a negative feedback system that keeps the temperature within a set range. Imagine a "reverse" thermostat that provides positive feedback such that hotter the building becomes, the more it increases the heating. This system is unpredictable, but not likely to last long. Thus, the systems of interest are those in the middle of the spectrum. Chaotic systems are now widely familiar. They result from

having a very high sensitivity to initial conditions and their behavior fluctuates wildly, but within a predictable *range* of states.

More interesting are systems at the edge of chaos. Here the system has the freedom to explore new possibilities and may "jump" to a new and higher form of organization. They are characterized by goal-directedness, at least insofar as they operate in order to maintain themselves. In the process of self-maintenance they create their own components. Evan Thompson says that a living cell is a paradigm case. Its constitutive processes are chemical; "their recursive interdependence takes the form of a self-producing, metabolic network that also produces its own membrane; and this network constitutes the system as a unity in the biochemical domain and determines a domain of possible interactions with the environment" (44).

We reach a new level of complexity in systems that operate not on the basis of predetermined goals and feedback loops (for example, the homeostatic systems in an organism) but also have the capacity to select their own goals, and thereby adapt to new circumstances. These are called complex *adaptive* systems. When such systems also have some sort of memory, a way of storing information about what has or has not worked in the past (for example, the genome) there is heightened ability for the system (here a species) to increase its adaptation over time. The capacity for memory in individual organisms brings us to the point of being able to speak of information and meaning. This adaptive selection opens the possibility of learning and the emergence of novel behavior, based in neural plasticity and the ongoing influence of events outside of the organism.

Complex adaptive systems theory has dramatic consequences for understanding causation. While ordinary efficient causation is presupposed, systems theory developed specifically because such causation is inadequate to describe complex systems. This is in part because complex systems operate on information as much as on energy and matter. More important is the fact that the relations among the components of a system need to be thought of in terms of *constraints*. An efficient cause makes something happen. A constraint *reduces* the number of things that can happen, as a result of the fact that the components are internally related to one another. Thus, a change in one automatically changes the other. Juarrero says: The concept of a constraint in science suggests "not an external force that pushes, but a thing's connections to something else . . . as well as to the setting in which the object is situated" (1999, 132). More generally, then, constraints pertain to an object's connection with the environment or its embeddedness in that environment. They are relational properties rather than primary qualities in the object itself. Objects in aggregates do not have constraints; constraints only exist when an object is part of a unified system.

From information theory Juarrero employs a distinction between context-free and context-sensitive constraints. For example, in successive throws of a die, the numbers that have come up previously do not constrain the probabilities for the current throw; the constraints on the die's behavior are context-free. In contrast, in a card game the constraints are context-sensitive: the chances of, say, drawing an ace at any point in the game are sensitive to history because the rules of the game, the number of cards in the deck, and so forth, create relations among the possible outcomes such that the probability of one occurrence is related to all of the others. This account suggests that a better term

in place of “downward causation” is “whole-part constraint.” The “higher-level” system, the whole, does not exert efficient, forceful causation on its components. Rather, global features of the system are such that a change in one component changes the probabilities of the occurrence of other lower-level events. So here is an explanation of Van Gulick’s selection.

Due to the role of probability in complex systems, it is necessary to do away with the sharp distinction between determinism on the one hand, and indeterminism (that is, quantum indeterminacy or complete randomness) on the other. The appropriate middle term is “propensity,” coined by Karl Popper to mean “an irregular or non-necessitating causal disposition of an object or system to produce some result or effect” (Sapire 1995, 657, referring to Popper 1990).

An understanding of the concept of a propensity has been aided by the study of nonlinear mathematics and especially chaotic systems. It begins with a visual or imaginary “state space” or “phase space,” which is an *n*-dimensional space. In this space, a trajectory represents possible transitions from one state of the system to another. Chaotic systems theory introduced the concept of a “strange attractor” to describe the development of chaotic systems over time. This is a “shape” in phase space that depicts the boundaries within which the system can be found during its evolution.

From the concept of a strange attractor the idea of an “ontogenic landscape” has been developed. This is a “topographical map” in which valleys represent areas in phase space in which the system is likely to stay. Peaks represent states in which the system will only be found as a result of a major perturbation, such as the injection of a great deal of energy. So the system has a propensity to remain within the valleys. The topography

represents a summation of the general effects of a vast number of contextually constrained interactions among the system's component processes.

In sum, complex adaptive systems theory postulates that such systems become causal players in their own right, partly independent of the behavior of their components, selectively influenced by the environment, and capable of pursuing their own goals.

Sec. 5: An Example

To illustrate how the concepts drawn from complex system theory operate, Brown and I chose the "simplest" complex system we could imagine, an ant colony. Harvester ant colonies consist of a queen surrounded by interior workers deep inside the burrow, and other worker ants that only enter chambers near the surface. The worker ants are specialized: some forage for food, others carry away trash, and still others carry dead ants away from the colony. Ant colonies show various sorts of 'intelligent' behavior. If the colony is disturbed, workers near the queen will carry her down an escape hatch. "A harvester ant colony in the field will not only ascertain the shortest distance to a food source, it will also prioritize food sources, based on their distance and ease of access. In response to changing external conditions, worker ants switch from nest-building to foraging, to raising ant pupae."[1] The means by which this happens is based on "ant rules" such as "if a forager ant crosses more than three pheromone trails from other foragers within a minute she returns to the nest."

Colonies develop over time. Successful colonies last up to fifteen years, the lifespan of the queen, even though worker ants live only a year. The colonies themselves

[1] Gordon, quoted in Seven Johnson, *Emergence: The Connected Lives of Ants, Brains, Cities, and Software* (New York: Scribner, 2001), 84,

go through stages: young colonies are more fickle than older ones. Younger colonies are also more aggressive. "If older colonies meet a neighbor one day, the next day they're more likely to turn and go in the other direction to avoid each other. The younger colonies are much more persistent and aggressive, even though they're smaller."[2]

It is tempting to try to explain the behavior of the colony reductionistically. Knowledge of some of the "ant rules" gives the impression that the behavior of the colony is entirely determined bottom-up. One can imagine that each ant has built-in laws governing its behavior, and one can imagine a molecular-neural level account: "smell of fourth forager within one minute causes return to the nest." So the typical causal agent is not "the system as a whole" or "the environment" but a few molecules of a pheromone embedded in the ant's receptor system. If one had all of the information about the rules, the initial placement of the ants, and the pheromone trails one could predict or explain the behavior of the whole colony.

Now consider an alternative, systems-theory description of the phenomena. The colony as a whole is certainly describable as a system. It is bounded but not closed; it is a self-sustaining pattern. The shift in perspective required by a systems approach is to see the colony's components as a set of interrelated *functional* systems—not a queen plus other *ants,* but rather an *organization of processes* such as reproduction, foraging, nest-building. It is a self-organized system that runs on information; it produces and maintains its own functional systems in that the relations among the ants constrain them to fulfill the roles of forager, nest-builder, etc. In addition it has a high degree of autonomy vis-à-vis the environment.

---

[2] Ibid.

The colony displays a number of emergent, holistic properties. In addition to its relative stability there is the "intelligence" displayed in the placement of the trash pile and cemetery, the ability to prioritize food sources. Accidents of the environment such as location of food sources affect the foraging system as a whole, which in turn constrains the behavior of individual ants.

The crucial shift in perspective is from thinking in terms of causes (that is, nothing will happen unless something makes it happen) to thinking in terms of both bottom-up causes *and* constraints (that is, a variety of behaviors are possible and the important question is what constricts the possibilities to give the observed result). It is a switch from viewing matter as inherently passive to viewing it (at least the complex systems in question) as inherently active. In contrast to the *assumption* that each lower-level entity will do only one thing, the assumption here is that each lower-level entity has a repertoire of behaviors, one of which will be *selected* due to its *relations* to the rest of the system and to its environment. Ants in the colony respond to context-sensitive constraints that entrain their behavior to that of other ants in ways sensitive to history and to higher levels of organized context.

Sec. 6: From Ants to Moral Responsibility

It's important to point out that the level of complexity involved in an ant colony is comparable to that of the very simplest of multi-celled organisms—those without a nervous system. To get from ants to human conscious choices it is necessary first to consider the ways in which all complex organisms differ from simple ones. The variables that lead to increases in the capacity for self-causation include modifiability of

parts, neural complexity, behavioral flexibility, and increasing ability to acquire information. In systems terms, this involves functional specialization of components and a high level of flexible coupling of those components. As we move from rudimentary animal behavior toward humans, we see a vast increase in brain size, tighter coupling in terms of number of axons, dendrites, and synapses, structural complexification, recurrent neural interconnections, and complex functional networks that are hypothesized to be the source of consciousness.

But still there is the question of what distinguishes intelligent, self-conscious, and morally responsible choice from the flexibility and autonomy of the other higher animals. Brown and I argue that the two crucial developments are symbolic language and the related capacity to evaluate one's own behavior and cognition. Thus, the remainder of the story about how we get to moral responsibility and some form of free will involves, first, consideration of the charge that brain science cannot make sense of linguistic meaning. We argue that the supposed mysteries of meaning and intentionality are a product of left-over Cartesian assumptions regarding the inwardness of mental acts and the passivity of the knower. If instead we consider the mental in terms of action in the social world, there is no more mystery to how the word 'chair' hooks onto the world than there is to how one learns to sit in one. We consider what is known so far about the neural capacities needed for increasingly complex use of symbols. Symbolic language—in fact, quite sophisticated symbolic language—is a prerequisite for both reasoning and morally responsible action.

We then turn to the question: "How does reason get its grip on the brain?"—that is we turn to the role of reason in human thought and action. A powerful reductionist

argument is the lack, so far, of a suitable account of “mental causation,” that is, of the role of reason in brain processes. The problem is often formulated as the question of how the mental properties of brain events can be causally efficacious. We reformulate the problem, instead, as two questions: how is it that series of mental/neural events come to conform to rational (as opposed to *merely* causal) patterns; and what difference does the possession of mental capacities make to the causal efficacy of an organism's interaction with its environment?

These moves put us in position to consider the central theme of our book, a philosophical analysis of the concept of morally responsible action. Here we adopt an account of moral agency worked out by Alasdair MacIntyre. Morally responsible action depends (initially) on the ability to evaluate that which moves one to act in light of some concept of the good. We then investigate the cognitive prerequisites for such action, among which we include a sense of self, the ability to predict and represent the future, and high-order symbolic language.

Finally, we bring to bear our argument to the effect that organisms are (often) the causes of their own behavior, together with our work on language, rationality, and responsibility, in order to make the claim to have eliminated one of the worries that seems to threaten our conception of ourselves as free agents, namely neurobiological reductionism—the worry that "my neurons made me do it."

Conclusion

The purpose of this paper has been to answer as well as possible in such a brief time the question of why it is not the case that, in normal circumstances, we can use neurobiology

to relieve us of moral responsibility. I have claimed that the key to preserving traditional concepts of morality is the defeat of neurobiological reductionism. The main emphasis of my paper was on the way complex adaptive systems theory lays the conceptual groundwork for defeat of reductionism in general, and especially of neurobiological reductionism.

I have reported on current accounts of the ways in which complex adaptive systems, as they increase in complexity, become more and more autonomous. They take partial control over their own components; they interact selectively with their environments; they pursue their own goals. In short, they are *agents*. I have very briefly sketched the path Brown and I have followed to move from the *absence* of determinism in complex organisms an account of moral responsibility and free will. All of this has been in the service of the theological motive of explaining why physicalism is an acceptable position for Christians.