ISSN 2045-5577

ORDER: GOD'S, MAN'S AND NATURE'S

Avoiding Neurobiological Reductionism: The Role of Downward Causation in Complex Systems

by Nancey Murphy

(This chapter will appear in Moral Behavior and Free Will. A Neurological and Philosophical Approach (eds *Juan José Sanguineti, Ariberto Acerbi, José Angel Lombo*)

1 Introduction

I am pleased that my essay follows that of Carlos Moya in this volume. His chapter explains clearly the problem of neurobiological reductionism, in that if reductionism is true, then there is no way to make sense of mental causation. Furthermore, if we cannot make sense of mental causation, then neither can we account for moral responsibility or free will—how can a person's actions be attributed to his or her intentions? In fact, if there is no such thing as mental causation, then there is no point in attempting to write arguments about reductionism, since there is no way to understand how an argument could be constructed or accepted on the basis of reason. If neurobiolgical reductionism is true, there can be no such thing as an argument for it!

Thus, my plan is to provide insights, largely from philosophy of science, to criticize the reductionism that usually plagues discussions of relations between the neural and the mental in both neuroscience and philosophy of mind. First I distinguish several different reductionist theses, and note that causal reduction will be the focus of this study. Next I present a brief history of attempts to avoid causal reductionism, beginning in the early twentieth century, and incorporating the best of current thinking on emergence and downward causation. These recent developments provide nearly all of the insights needed for what Alwyn Scott calls a paradigm change across the sciences—the development, over the past generation, of systems theory. 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. After describing and illustrating systems-theoretical concepts, I argue that this development is just what is needed to solve the problem of mental causation. I also sketch in the briefest way what my co-author Warren Brown and I see as the path from an understanding of mental causation to accounts of moral responsibility and free will (Murphy and Brown 2007).

2 Varieties of reductionism

There are a number of related reductionist theses. Their relations can be understood in terms of the assumption, as Moya says, of "a layered and hierarchical view of the world in which the physical level is metaphysically basic" (typescript p. 4). So there are two (idealized) hierarchies. In the model of (1) the hierarchy of the *sciences*, the sciences are taken to pertain to (2) a hierarchy of increasingly *complex systems,* each system composed of the entities from the next level down.

Methodological reductionism is the research strategy of analyzing the thing to be studied into its parts. This was *the* focus of early modern science, but in many fields it is now recognized to be inadequate. It needs to be complemented by attention to the environment of the system under study. This point is central to Moya's work on mental causation, as well as to Brown's and my conclusion that the mental cannot be understood entirely in terms of the neural substrate, but only by considering, in addition, the contextualization of the brain in the whole body, and in the person's behavior, history, and natural and socio-cultural environments.

Epistemological or theoretical reductionism is the view that laws or theories pertaining to higher levels of the hierarchy of the sciences can (and should) be shown to follow from lower-level laws, and ultimately from physics. This was the goal of the neo-positivists, such as Ernst Nagel, referred to in Moya's discussion of Jaegwon Kim's philosophy of mind.

A number of authors also distinguish *ontological* reductionism, but this term is used for two different positions. One is the unobjectionable thesis that as one goes up the hierarchy of complex systems, no new ontological "ingredients" need to be added—for example, no vital force is needed to get life from inorganic material, no mind or soul is needed to get sentience or consciousness. The stronger (and highly objectionable) thesis is that only the entities at the lowest level of the hierarchy are "*really* real";

molecules, cells, organisms, are merely temporary aggregates of the elementary entities.

Causal reductionism follows rather directly from strong ontological reductionism. This is the view that behavior of the parts of a system (ultimately, the parts studied by subatomic physics) is determinative of the behavior of all higher-level entities. Thus, this is the thesis that all causation in the hierarchy is "bottom-up."

The major threat to understanding the relations between the mental and the neural is causal reductionism. This is the assumption that human beliefs and behavior are determined "bottom up" by neural processes. That is, parts (here neurological and other biological parts) unilaterally determine thought and behavior, and thus genuine free will and moral responsibility must be illusory. This is also the source of the problem of mental causation.

3 Responses to causal reductionism

The most significant criticisms of causal reductionism (and thereby of methodological and epistemological reductionism) fall into three historical stages: an early emergentist movement (from approximately 1920-1950); the exploration of the concept of downward causation beginning in the 1970s; and, currently, new attempts to define and defend both emergence and downward causation).

3.1 The early emergentist movement

The idea of emergence was proposed in philosophy of biology as an alternative both to mechanist-reductionist accounts of the origin of life and to vitalism. Vitalists claimed that in order to get life from inorganic matter something like a vital force needed to be involved. Emergentists such as Roy Wood Sellars argued that the increasingly complex organization, as one ascends the hierarchy of systems, accounts for the appearance of new kinds of entities with causal capacities that cannot be reduced to physics. The organic emerges from the physical; so too do the levels of the mental or conscious, the social, the ethical, and the religious or spiritual ([1932] 1996; 1970).

Sellars claimed that "the ontological imagination" was stultified by the picture of microscopic billiard balls. In contrast, he argued that organization and wholes are genuinely significant; they are not mere aggregates of elementary particles. Reductive materialism overemphasizes the "stuff" in contrast to the organization. "There is energy; there is the fact of pattern; there are all sorts of intimate relations." "Matter, or stuff,

needs to be supplemented by terms like integration, pattern, function" (1970, 136-38 *passim*). With hindsight we can see that Sellars and some of the other emergentists were exactly right; however, their arguments did not prevail against the reductionist philosophers of science.

3.2 First sketches of downward causation

In the 1970s psychologist Roger Sperry and philosopher Donald Campbell both wrote about downward (or top-down) causation. Sperry sometimes wrote of the higher-level entity or system *overpowering* the causal forces of the component entities (1983, 117); this rightly raised worries regarding the compatibility of his account with respect for the basic sciences.

Campbell does not write of overpowering lower-level processes, but instead provides a non-mysterious 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. He intends to illustrate four theses; the first two give due recognition to bottom-up accounts of causation: First, all processes at the higher levels are restrained by and act in conformity to the laws of lower levels, including physics. Second, the achievements at higher levels require for their implementation specific lower-level mechanisms and processes. Explanation is not complete until these micromechanisms have been specified.

The third and fourth theses represent the perspective of downward causation: Third, "[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." Fourth:

> 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. (1974, 180)

Unfortunately, it appears that little was written on downward causation in philosophy until the 1990s.

3.3 The current status of downward causation

The concept of downward causation appeared in philosophy of mind in the 1990s, but it has been given short shrift by the very influential Jaegwon Kim. Richard Rorty wrote that it is pictures rather than propositions that determine most of our philosophical convictions (1979, 12). Discussions of reductionism in philosophy of mind have been stultified by an unhelpful picture—a major obstacle to understanding downward causation in that sphere (Murphy and Brown 2007, 233-36). Psychology is above biology in the hierarchy of the sciences, so mental properties are higher-level properties than brain properties. Thus, it is possible to create a picture such as the following, in which *M* and *M** represent mental properties; *B* and *B** represent brain properties; and the horizontal arrow represents the presumed causal relation between the brain properties. The vertical lines usually represent the (now fashionable) *supervenience* relation (see Moya 2010, ts, p. 11).[1]

```
M              M*
|              |
B  -------->   B*
```

With this picture available, philosophers of mind ask what causal work is left for *M*. Here is Kim's tri-lemma: One possibility is to give up on *any* causal role for the mental—the mental is epiphenomenal—and this amounts to reductive physicalism. Another is to count both *B* and *M* as sufficient causes of *B** (and *M**) but this would amount to causal overdetermination, which Kim simply dismisses as "absurd."

Finally, Kim considers that there may be some form of downward causation from *M* to *B** (i.e., *M* and *B* are the jointly sufficient cause of *B**), but he rejects this possibility because it "breaches the causal closure of the physical domain." And ". . . to give up this principle is to acknowledge that there can in principle be no complete physical theory of physical phenomena, that theoretical physics, insofar as it aspires to be a complete

[1] See Murphy and Brown (2007, 205-09) for objections to the now-predominant definitions of supervenience, which, we argue, entail reducibility. It needs instead to be defined so as to incorporate the role of the *context* of brain events.

theory, must cease to be pure physics and invoke irreducibly non-physical causal powers—vital principles, entelechies, psychic energies, elan vital, or whatnot" (1995, 208-09). Thus, Kim opts for reductive physicalism, claiming that mental properties are causally efficacious only to the extent that they are reducible to physical properties.

A much more helpful move comes from philosophy of science. Robert Van Gulick 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 (cf. Sellars, sec. 3.1); 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 what might be called downward causal influences without requiring any objectionable form of emergentism by which higher-order properties would alter the underlying laws of physics. Higher-order properties act by the *selective activation* of physical powers and not by their *alteration*" (1995, 252).

From Van Gulick's account we can see that evading causal reductionism will require the recognition that higher-level entities and systems have emerged (evolved) from lower, and that these entities can be somewhat independent of the causal processes of their constituents, thereby manifesting new, higher-level causal capacities.

3.4 New defenses of emergence

I noted above that the early emergentist movement was overwhelmed by the reductionist program of the neopositivists. And while we can see that thinkers such as Sellars were on the right track, the movement perhaps needed to die out given the variety and vagueness of many of its theories. New attempts have been made recently to clarify the meaning of "emergence." Van Gulick, again, is a valuable resource (2001). He argues that one can create a typology of emergentist theses by taking them to be *denials* of the various and well-defined reductionist theses (some of which are catalogued above). If *causal* reductionism is important for philosophy of mind, then a theory of the emergence of higher-level causal capacities will be most relevant here.

I believe that the best account of emergence so far is that of Terrence Deacon (2007). Deacon distinguishes three types or levels of emergence. There is no emergence in mere aggregates, though an aggregate does have one sort of global properties. For example, the weight of a volume of liquid is a simple addition of the weights of its molecules.

The important difference between an aggregate and a system is that in a system it is *relational* properties of the constituents (as opposed to primary or intrinsic properties) that constitute the higher order. In such cases additional configurational and distributional information is needed to account for the higher-order properties. Deacon includes here the viscosity of liquids, turbulence in large bodies of water, and typical feedback systems such as a thermostatically controlled heating system. He calls this first-order emergence. Because fluctuations in such systems are dampened out across time it is possible to give (rough) reductionist accounts of their behavior.

Second-order emergence occurs when there is symmetry breaking or the *amplification* of a fluctuation rather than dampening. Systems in which this occurs are nonlinear; their history matters. There are simpler and more complex versions of such systems. The simpler sort is self-organizing, in that higher-order patterns selectively constrain the incorporation of lower-order constituents into the system or select among possible states of the lower-level entities (Van Gulick's point, as well). More complex

second-order emergent systems are also autopoietic: they change the lower-order constituents themselves. Examples of the simpler sort are the Bénard phenomenon (the development of orderly convection rolls in a heated liquid), a thermostat that amplifies rather than dampens feedback, and the development of a snowflake. An autocatalytic cycle is of the more complex sort in that the system manufactures some of its own components. All life involves second-order emergence of the more complex sort.

Deacon distinguishes between first- and second-order (as well as third-order) emergence in terms of what he calls "amplification logic" or "the topology" of causal processes. In systems without emergence, global properties are all produced bottom-up (or by means of *local* interactions with boundaries—e.g., a water molecule constrained by the presence of the surface of the container). In first-order emergent systems there is "nonrecurrent" causal architecture: a simple bottom-up and top-down relation in which global properties of the system (e.g., density of components) makes a difference to the relations among components and thus to the behavior of the whole system.

Second-order systems have more "tangled" or "recurrent" causal architecture as a result of the amplification of lower-level fluctuations. This amplification changes the total state of the system in a way that makes a decisive difference for the future development of the system. This can lead to new orders of complexity.

Third-order emergence involves the interaction among three levels and appears (naturally) only in the biological realm. Here a variety of second-order forms emerge, and are selected (constrained) by the environment, but in such a way that a *representation* of its form is introduced into the next generation—the system has "memory." The simplest example is the evolutionary process. The micro-level (the genome) in interaction with the organism's environment, directs the construction of the organism (the mid-level), whose reproductive fate is determined top-down by the environment (top level). The preservation of information regarding the organism's success in the environment is the means by which relatively stable populations of successful organisms can be produced, within which future fluctuations appear. Some of these may be amplified (preserved and re-entered into the system) by means of interaction with the environment, thus enabling the appearance of still higher degrees of complexity. Deacon describes such systems as exhibiting recurrent-recurrent causal architecture: over time, a two-stage process of emergence occurs that results in downward causation not just from top to mid-level, but from top to bottom (environment to genome).

Deacon's account makes it obvious that downward causation and emergence are complementary concepts; that it, the evolutionary process serves as an illustration of emergence for Deacon and of downward causation for Campbell. The entanglement of bottom-up and top-down causal processes is what makes causal reduction impossible. Furthermore, Deacon's third-order emergence describes the processes by which more complex, higher-order systems come into existence.

4 Mental causation: Taking stock

I began section 3.3 with the typical visual representation of the problem of mental causation, and Kim's judgment that downward causation provides no solution. I also mentioned my conclusion (with Brown) that this simple drawing is one of the major *obstacles* to solving the problem of neurobiological reductionism. One reason is based on Deacon's tri-level account of emergence: the relations of mental to physical phenomena cannot be understood in terms of first-order, supervenient emergence. Third-order (evolutionary) emergence contains second-order (self-organizing) emergence as a limiting case, which in turn contains first-order (supervenient) emergence as a limiting case. For this reason it is mistaken to represent a single mental phenomenon as directly supervenient on a single brain event. One needs to appreciate the many levels of embedded evolutionary emergent processes characteristic of brains, as well as the fact that typical mental events can only have meaning insofar as they are contextualized within complex systems of mental/brain events.

Moya's argument for mental causation provides a second critique of typical work in philosophy of mind. He emphasizes the roles of action and socialization, drawing on the theory of "neural Darwinism" (although without using the term). Neural connections form spontaneously (bottom-up causation). The connections that subserve some "fitting" mental operation, and are therefore reinforced by the environment, are strengthened; others weaken or die off. This is the neural basis for learning. For example, random neural "noise" produces childhood babbling. The parents' *response* to the random production of, say, "ma ma" acts as a downward cause from the environment to the child's behavior, and thereby strengthens the relevant neural connections. At a more advanced level, cultural semantic and other systems act as downward causes to structure human brains *and bodies* so as to think and act rationally. Moya emphasizes the role of the entire body, acting in physical and social environments, because it is only by means of action that the feedback needed for learning can take place.

Moya's account is both similar and complementary to the account of mental causation developed by Brown and myself. We also reject what we call "Cartesian materialism"—the substitution of inner *brain* processes for Descartes's inner mental processes, thereby ignoring the role of the whole body in its environment. [2] Brown argues that mental events need to be understood as *contextualized* brain events. Mental events are typically contextualized in action-feedback-evaluation-action loops in the environment.

Consider again the example of a baby learning to speak. Whatever the neural processes (brain events) that normally underlie the toddler's ability to utter "ma ma," these processes could mean nothing without, first, the bodily capacity for verbalization. Second, the verbalization means nothing apart from a linguistic system (ongoing context) in which the sound "ma ma" can be interpreted as a word referring to the child's mother. Temporal contexts are relevant as well. If the child never goes on to use the word we would withdraw our judgment that it had learned the word at all. Past context matters as well; the history of reinforcement is crucial. If a child in a very different linguistic community just happens utter "ma ma" in the presence of its mother we are not entitled to say that it has learned the English word "mamma."

Are *all* mental events necessarily contextualized? Probably not. Newborn children experience pain, presumably of a contextless sort. But mature human pain is partially constituted by context. Small children soon learn to direct their natural cries of pain toward adults in order to get attention. As Ludwig Wittgenstein pointed out, they later learn to substitute language for the natural cries. Later still, the role of pain as a signal of bodily damage is learned and so pain now *means* something that it could not before, and it now serves as a motive for future action—to find a diagnosis and a cure.[3]

Is the problem of mental causation now solved by exchanging Deacon's complex account of third-order emergence for the simple supervenience relation employed in much of philosophy of mind, and by invoking downward (selective) causation from environment to individual, and thence to neural connections and events? We can certainly see that Kim's description of downward causation is off the mark. We are not postulating that the "higher-level" mental events possess some sort of "irreducibly non-

---

[2] Daniel Dennett coined the term, but we use it in a somewhat different sense, and accuse him of being one of the worst offenders (see Murphy and Brown 2007, 29, 295-96).

[3] I thank Uwe Meixner for calling my attention (during the conference) to this possible counter-example.

physical causal powers" such as "psychic energies." Rather, the mental level is "higher" in the sense that it is part of a higher-level (broader) system, and that system has effects (over time) at the neural level.

However, there are still legitimate worries. The objection that can be raised against an account of downward causation based on selection among of lower-level causal capacities is the following: If the higher-order pattern or system is *selecting* among the causal capacities of its constituents, then this amounts to *causing* one such capacity to be activated rather than another. How can we make sense of *this* causation? If it is not by means of something like Kim's psychic forces, then it must be by means of the same sorts of causal forces that are recognized at the level of physics. If so, then bottom-up causation is doing all of the work after all, and causal reductionism is true. A second and related problem is to explain how any such (apparent) downward causation could fail to amount to causal overdetermination.

These objections have led me to conclude that the sort of organization and selection of lower-level causal processes that Van Gulick describes call for new a set of concepts, and the concepts in question are constitutive of a shift from thinking in "mechanistic" terms to thinking in "systems" terms.

5 Complex adaptive systems theory

"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. Another early source was the study of cybernetics. 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. Mathematician and neuroscientist Alwyn Scott makes the bold claim that systems thinking represents a paradigm change across all of the sciences (2004, 2). 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.

5.1 The basics of systems theory

Several authors call for what might be called a shift in ontological emphases. Alicia 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). Francis 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 components of complex systems are not 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 *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. For example, a thermostatically controlled heating system is 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 the colder or hotter the building becomes, the more it increases the cooling or 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 familiar.

They result from having a sensitivity to initial conditions that falls into a narrow range and their behavior falls into a predictable *range* of states.

More interesting are those 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 (cf. Deacon's account of third-order emergence, sec. 3.4). An understanding of how this can happen in terms of physics comes from the study of far-from-equilibrium thermodynamics. Such systems are called complex *adaptive* systems.[4] They are characterized by goal-directedness, at least insofar as they operate in order to maintain themselves. In the process of self-maintenance they may create their own components. For example, in an autocatalytic reaction, molecule A catalyzes molecule B, which catalyzes more of A. The process will stabilize at some point unless additional materials are introduced into the system. In order for the system to maintain itself, the internal dynamics must determine which molecules are fit to be imported into the system and survive (Juarrero 1999, 126).

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 such that 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

[4] Other terms applied to such systems are "self-organizing," "autopoietic," and "dynamical."

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 (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 and indeterminism (quantum indeterminacy or complete randomness). 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 which 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.

Brown and I argue that this set of new concepts, particularly that of context-sensitive constraints, give us the conceptual tools to explain how downward "causes" cause without violating the causal closure of the physical and without postulating causal overdetermination.

5.2 A "simple" complex system

Brown and I chose ant colonies to illustrate the applicability of systems-theoretical concepts. While our interest is in humans as complex systems, they are much *too* complex to use as an illustration.

Harvester ant colonies consist of a queen and worker ants. Worker ants are specialized: some forage for food, others carry away trash, and still others carry dead ants away from the colony. These latter specialists manage to locate the trash pile and the cemetery at points that maximize the distances between cemetery and trash pile, and between both of these and the colony itself.[5]

Ant colonies show other sorts of "intelligent" behavior: If the colony is disturbed, workers near the queen will carry her down an escape hatch. An ant colony in the field ascertains the shortest distance to a food source and prioritize food sources based on their ease of access. In response to changing external conditions, worker ants switch from one specialization to another. The ability of ants to shift functions is explained by means of chemical signals called pheromones. The colony's ability to adjust task allocation according to colony size and food supply depends on what Deborah Gordon calls their ability to "conduct a kind of statistical sample of the overall population size, based on their random encounters with other ants." For example, "[a] foraging ant might expect to meet three other foragers per minute—if she encounters more than three, she might follow a rule that has her return to the nest" (Johnson 2001, 76-7).

Colonies develop over time, lasting up to fifteen years, the lifespan of the queen, although workers live only a year. They also go through stages: young colonies are more fickle and aggressive than older ones.

The colony illustrates the following attributes of a complex adaptive system: It is bounded but not closed; it is a self-sustaining pattern; it has a high degree of autonomy vis-à-vis the environment. The colony displays emergent, holistic properties: its relative stability, the "intelligence" displayed in the placement of the trash pile and cemetery, and the ability to prioritize food sources. It is a self-organized system that runs on information. Changes in the colonies' "attitudes" over time from fickle and aggressive to greater stability reflects the tendency of all complex systems to settle into more stable patterns.

---

[5] This account is taken from Steven Johnson (2001), whose work is based largely on that of Deborah Gordon (1999).

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 along with other *ants,* but rather an *organization of processes* such as reproduction, foraging, nest-building. 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, and so on. These are context-sensitive constraints that entrain ants' behavior to that of other ants in ways sensitive to history and to higher levels of organized context. From this point of view, the genetically imprinted rules in the individual ants' nervous systems are not to be understood as causal laws; they are receptors of information regarding such things as the density of the forager population. A holistic property of the system—forager density—increases the probability that a given forager will encounter more than three other foragers per minute, and thus the probability that the ant will return to the nest. It is a non-forceful constraint on the ant's behavior.

The reductionist assumption is that if one could know the placement of all of the ants, the pheromone trails, and the wiring of the ants' nervous systems, the behavior of the colony would be predictable. At this point, such a position must be seen merely as an assumption. However, even if this were true, the reductionist cannot explain, given that no two complex systems (e.g., two ant colonies) are ever identical, why it is the case that, starting from wide varieties of initial conditions, one finds such similar patterns emerging. That the world is full of such phenomena is now a widely recognized fact, but it is counter-intuitive on a bottom-up account. The *fact* of higher-order patternedness in nature, patterns that are stable despite perturbations, and despite replacement of their constituents, calls for a paradigm shift in our perceptions of (much of) the world.

6 Mental causation, morality, and free will

Moya notes that Van Gulick's proposal "applies especially well to cases in which, say, biological patterns activate underlying physical powers," but he is skeptical that it is sufficient to understand how downward causation functions in the case of intentional patterns (2010, ts. p. 15). In our work, Brown and I first thought that the defeat of neurobiological reductionism by means of an account of downward causation would be sufficient to account for mental causation and thus dissolve threats to moral responsibility and free will. Reading Juarrero's work (1999), however, convinced us that we still had far to go in order to give an account of how humans can be the causes of their own behavior (see also Juarrero 2009).

Complex adaptive systems theory postulates (and I hope that my ant colony illustrates) 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. However, 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 (number of axons, dendrites, 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, we consider the nature of meaning and intentionality, and, as does Moya, we approach meaning in terms of action in the social world. Symbolic language—in fact, quite sophisticated symbolic language—is a prerequisite for both reasoning and morally responsible action. We then turn to the problem of mental causation.

Moya's approach to mental causation relies on Fred Dretske's distinction between structuring and triggering causes. Moya claims, with Dretske, that intentions serve as structuring causes, but he also agrees with Terrence Horgan in arguing that reasons or intentions should be seen as full-fledged triggering causes of action as well. I believe that systems theory has an important role to play here. Juarrero makes use of this theory to answer the question of how intentions guide behavior. She would agree in calling an intention a triggering cause. However, her interest is in explaining how the content of an intention "can inform and flow into behavior such that the action actualizes the content of that intention" (1999, 5). For this, she argues, we need to understand the brain's self-organization in terms of complex adaptive systems theory, and an intended act type as a strange attractor, guiding action over time through the contingencies and

perturbations that would otherwise tend to derail it. The formation of an intention alters the ontogenic landscape representing one's awareness of future actions in such a way as to guide the action to completion.[6]

Juarrero's account of intentions as attractors would apply to both humans and other higher animals. So a remaining question is what distinguishes adult humans' morally responsible actions from those of animals and even small children. Brown and I adopt an account of moral agency worked out by Alasdair MacIntyre (1999). Morally responsible action depends (initially) on the ability to evaluate one's reasons for acting in light of a 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—the argument I have made briefly in this paper—together with our work on language, rationality, and responsibility, in order to make the claim to have eliminated one of the worries that seem to threaten our conception of ourselves as free agents, namely neurobiological reductionism—the worry that "my neurons made me do it."

6 Conclusion

The purpose of the present volume is to examine the relationship of the philosophical (and theological) concepts of moral responsibility and free will to current work in the cognitive neurosciences. I have claimed in this chapter that the key to preserving traditional concepts of morality and freedom is the defeat of neurobiological reductionism; 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, dynamical, autopoietic) 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—at least of self-maintenance. In short, complex adaptive systems are (to a greater or lesser extent) *agents*. I have briefly described how Juarrero uses the concepts of strange attractors and ontogenic landscapes to reflect on the role of intentions not only in triggering actions but in guiding them to completion. I have also very briefly sketched the path Brown and I

[6] Due to space constraints I shall not attempt to summarize her application of systems theory to the problem of free will.

have followed to move from the *absence* of neurobiological determinism to defeat one of the most pressing threats to moral responsibility and free will.

Bibliography

Campbell, D.T. 1974. "'Downward Causation' in Hierarchically Organised Biological Systems." In F.J. Ayala and T. Dobzhansky (eds.). *Studies in the Philosophy of Biology*, Berkeley and Los Angeles: University of California Press, pp. 179-186.

Deacon, T.W. 2007. "Three Levels of Emergent Phenomena." In Murphy and Stoeger, (eds), pp. 88-110.

Gordon, D. 1999. *Ants at Work: How an Insect Society is Organized.* New York: Free Press.

Heil, J. and A. Mele (eds). *Mental Causation.* Oxford: Clarendon, 1995.

Heylighen, F. 2011. "Self-Organization of Complex, Intelligent Systems: The ECCO Paradigm for Transdisciplinary Integration." *Integral Review* (forthcoming).

Johnson, S. 2001. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. New York: Scribner.

Juarrero, A. 1999. *Dynamics in Action: Intentional Action as a Complex System.* Cambridge, MA: MIT Press.

_______. 2009. "Top-Down Causation and Autonomy in Complex Systems." In Murphy et al. (eds.), pp. 83-102.

Kim, J. 1995. "The Non-Reductivist's Troubles with Mental Causation." In Heil and Mele (eds.), pp. 189-210.

MacIntyre, A. 1999. *Dependent Rational Animals: Why Human Beings Need the Virtues.* Chicago, IL: Open Court.

Moya, C.J. (201x). "Mind, Brain, and Downward Causation." In Sanguinetti et al. (eds.). *oral Behavior and Free Will: A Neurological and Philosophical Approach.* forthcoming.

Murphy, N. and W.S. Brown. 2007. *Did My Neurons Make Me Do It?: Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*. Oxford: Oxford University Press).

Murphy, N. and W.R. Stoeger, SJ. (eds.). 2007. *Evolution and Emergence: Systems, Organisms, Persons*, Oxford: Oxford University Press.

Murphy, N. et al. (eds). 2009. *Downward Causation and the Neurobiology of Free Will.* Berlin and Heidelberg: Springer Verlag.

Popper, K.R. 1990. *A World of Propensities.* Bristol: Thoemmes.

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Sapire, D. 1995. "Propensity." In R. Audi (ed.). *The Cambridge Dictionary of Philosophy.* Cambridge: Cambridge University Press, p. 657.

Scott, A. 2004. "A Brief History of Nonlinear Science." *Revista del Nuovo Cimento* 27:1 -115.

Sellars, R.W. 1932/1996. *The Philosophy of Physical Realism.* Reprint ed. New York: Russell and Russell.

_______. 1970. *Principles of Emergent Realism: The Philosophical Essays of Roy Wood Sellars*, W. P. Warren, (ed.). St. Louis, MO: Warren H. Green, Inc.

Sperry, R. W. 1983. *Science and Moral Priority: Merging Mind, Brain, and Human Values.* New York: Columbia University Press.

Van Gulick, R. 1995. "Who's in Charge Here? And Who's Doing All the Work?." In Heil and Mele (eds.), pp. 233-256; reprinted and condensed in Murphy and Stoeger (eds), pp. 74-87.

_______. 2001. "Reduction, Emergence, and the Mind/Body Problem: A Philosophic Overview. *Journal of Consciousness Studies* 8: 1-34; reprinted and condensed in Murphy and Stoeger (eds.), pp. 40-73.