The Journal of Mind and Behavior
Summer 2001, Volume 22, Number 3
Pages 311–334
ISSN 0271–0137

Can Dynamical Systems Explain Mental Causation?

Ralph D. Ellis

Clark Atlanta University

Dynamical systems promise to elucidate a notion of top–down causation without violating the causal closure of physical events. This approach is particularly useful for the problem of mental causation. Since dynamical systems seek out, appropriate, and replace physical substrata needed to continue their structural pattern, the system is autonomous with respect to its components, yet the components constitute closed causal chains. But how can systems have causal power over their substrates, if each component is sufficiently caused by other components? Suppose every causal relation requires background conditions, without which it is insufficient. The dynamical system is structured with a tendency to change background conditions for causal relations anytime needed substrates for the pattern's maintenance are missing; under the changed background conditions, alternative causal relations become sufficient to maintain the pattern. The system controls the background conditions under which one or another causal relation can subserve the system's overall pattern, while the components remain causally closed under their given background conditions.

In reconciling the phenomena of consciousness with those of neurophysiology, the problem of mental causation is one of the major stumbling blocks, especially for theories of the mind–body relation that hope to avoid a straightforward epiphenomenalism in which the apparent causal power of conscious intentions would be only an illusion. This paper examines a new possible solution to the problem offered by a dynamical systems approach to the theory of consciousness, considering some objections against it, and developing a way to respond to these objections based on a careful analysis of the way causal theory interrelates with the idea of a dynamical system.

Traditionally, philosophers have formulated the problem of mental causation in this way: the "causal closure of the physical realm" — the notion that every physical event, if it has a cause, must have a physical cause (Kim 1992, 1993) — seems empirically as true for physical and chemical events in the

Requests for reprints should be sent to Ralph Ellis, Ph.D., Department of Religion and Philosophy, Clark Atlanta University, Campus P.O. Box 81, 223 James P. Brawley Drive, S.W., Atlanta, Georgia 30314.

brain as it is elsewhere. Many empirical brain studies support this conclusion (Ellis and Newton, 1998a). Yet, when I raise my hand, it seems obvious that the *decision* to raise the hand makes the hand go up. If there is mental causation, this would mean that, if I make a mental choice or decision, the choice can have causal power over the resulting movements of my body. If the choice is merely an ephiphenomenon or causal side-effect of a series of micro-level events, over which "I" as a unified being have no control, then the feeling that the choice influences the resulting action would seem to be only an illusion. Both the feeling that the choice has causal power *and* the behavioral outcome are really caused by a sequence of micro-level events over which I have no control, except in the trivial sense that some of the previous micro-level events in my body control my behavior. This would be a trivial sense as far as "I" am concerned, in the same way that it would be trivial to say that, when a cancerous tumor grows in my brain, it is "I" who decides that it is to grow.

Thus, if the feeling that a mental decision has causal power is not to be a mere illusion, it is traditionally assumed that some combination of events other than the mental decision cannot be completely sufficient to produce the outcome independently of the decision. If A causes both B and C, then A is sufficient for C independently of B, so B is precluded from having any real causal power over C. B is simply an epiphenomenon of A. Because B always accompanies A, it *feels* as if B is causally necessary for C, but if the causal account is correct then this feeling must be only an illusion.

There are really two interrelated questions here, and much of their mystery stems from their interrelatedness. The question as to how a mental state can cause a physical event is only modestly mysterious if we assume a simple identity between certain combinations of mental and physical events. It becomes much more mysterious when we ask how a unified act of will on the part of a unified decision-making agent can cause the various micro-level physical events which all together comprise the physical substrata for this very agency itself. That is, how can the "self," simply in the sense of a unified being, voluntarily initiate the movements of its own bodily components? This question is mysterious for two reasons. Not only can a desire in the brain command the body to move, but also, and more paradoxically, a unified process called the "self" or "person" can command the components in the brain to conform to the patterns of activity needed to execute the desired thought processes, feeling states, perceptual imagery, and attentional activities. This notion that a whole being can somehow cause the movements of its own parts seems at odds with most scientific accounts of consciousness, in which the whole is built up *from* interacting components whose activities are caused in quite piecemeal fashion by various inputs, micro-mechanisms, and basic chemical reactions.

Holistic versus Bottom–Up Modes of Organization

It is tempting to just bite the bullet and opt for a solution which posits that the causal power of choice is an illusion. We could then simply posit that the movements are caused by interacting micro-level mechanisms, and leave the conscious choice out of the causal picture. The feeling that we choose to act would then be explained as a feeling that results from those same micro-mechanisms — an epiphenomenon of the events that exert the real causal power (Jackendoff, 1996; Searle, 1984; Smart, 1963, 1970).

Against this purely "bottom–up" solution, it can be argued that the feeling that our consciousness often does play an active role in organizing the micro-mechanisms leading to action is not a mere illusion. Voluntary movement really is different from an unconditioned reflex or an habituated neural-firing sequence (Jeannerod, 1997; Spence and Frith, 1999). In playing a well-practiced piece on the piano, for example, most of the movements are not consciously initiated: as soon as I tell my hands which piece to play, they automatically execute the sequence of notes, as a series of micro-mechanisms that require little direction except from the micro-mechanisms themselves, orchestrated mostly by the cerebellum at an unconscious level (Schmahmann, 1997). However, this is true only *until* I need to modify a learned sequence of motor commands. Then I must switch to the mode of voluntary, global directing of the micro-events (Jeannerod, 1997; Spence and Frith, 1999). I must consciously decide to hit the G before the A, and consciously command the selected finger to move.

There are both phenomenological and physiological differences between these two modes. It is not just an illusory difference. In the effortful movement, more widely distributed brain processes are quickly activated, and the whole pattern of the organization seems to be commanded, not just by the cerebellum, but in some sense by "me" as a whole. Brainstem arousal mechanisms activate neurotransmitters that permeate virtually all parts of the brain, and looping signals integrate the functions of the cerebellum, thalamus, hypothalamus, frontal and prefrontal areas and the anterior cingulate, which then lead to imagery involving parietal, occipital and temporal lobes, and tentative action imagery involving the motor cortex and supplementary motor area (Damasio, 1994, 1999; Jeannerod, 1997). In short, virtually the whole brain must be integrated at the point when a new decision is made, by contrast to the linear command sequences that are sent from the cerebellum when it is functioning on "automatic pilot," by means of conditioned responses and habitual motor programs.

In the deliberately self-controlled mode, a pianist can decide to use the fourth finger or the fifth, or play loud or soft, depending on the effect emotionally intended or wanted. Widely distributed aspects of the pianist's being participate in forming such motivations, as indicated by the analyses of the motivational brain systems by Panksepp (1998) and Watt (2000), and from this holistic formation flows the sequence of mechanisms needed to make the self-directed behavior happen. Widely distributed but holistically unified aspects of a golfer's being seem to execute the decision as to when the swing will start; there does not seem to be an isolated micro-mechanism that is responsible for the decision.

In the case of intentional choices, we seem to be in a very different realm from a simple linear mechanism, such as amygdala activation by the sudden sight of a snake (as in LeDoux, 1996). In that case, the standard explanation is quite simple, involving a very localized sequence of micro-mechanisms. In the case of forming the mood that is to be conveyed in music, the needed brain activity seems to be extensively global, simultaneous, and dependent on the precise timing of billions of micro-mechanisms occurring all through the brain all at once (Haines, Dietrichs, Mihailoff, and McDonald, 1997). The overall pattern, which is the mood, must in some way organize the activities of the micro-constituents. And this seems to be just the opposite of a situation in which the overall pattern is caused *by* those same micro-constituents.

Some will say that there is no paradox here, since any conscious state, whether global or localized, is just caused by the interaction of previous micro-mechanisms. While it is true that most scientific accounts of consciousness (perhaps excluding quantum brain theories) will regard every psychological state as having been caused by some combination of past events, this is not where the paradox lies. The paradox is how a holistic organizational pattern — once having been caused by whatever sequence of past events — can then have the power to organize its own micro-mechanisms so as to ensure that they will behave as dictated by the holistic organizational pattern. If an affective intention is to motivate behavior, then in order to maintain itself this affective state must use millions of simultaneous shunt mechanisms to ensure that just the right micro-events occur (Panksepp, 1998), and that just the right timing is maintained among them (Anderson, 2000), to allow the intended *quale* to continue being felt while at the same time all the micro-level events are being coordinated in such a way as to facilitate the flow of chosen actions from the affectively motivated choices of the unified self.

The problem of mental causation can thus be formulated in terms of the need to choose between two theoretical alternatives, each of which entails its own difficulties. Either (1) a whole organizational pattern has the power to control its own micro-constituents rather than only the other way around — which seems to fly in the face of causal closure at the micro-level (which in turn seems to be an empirical fact); or (2) the feeling that our choices determine our actions in a non-trivial sense is only an illusion.

The Illusory Choice Model: A Closer Look

A classic finding that is often interpreted as supporting the illusory-choice solution is that, prior to a subject's awareness of deciding to execute an action, the neurophysiological mechanisms that cause the action are already reliably measurable (Libet, 1999). A measurable readiness potential is observable .5 second before a willed action, whereas the subject is aware of the choice only .1 second before the act. Libet assumes that this means that the actual choice occurs unconsciously .4 second before we consciously will it. This assumes that the readiness potential is the correlate of the initiation of the action, and this seems to be the standard assumption among those who cite Libet's readiness potential as evidence that conscious choices are only causally irrelevant epiphenomena.

Actually, Libet tries to save "free will" from this problem by postulating that the conscious choice that occurs .4 second after the readiness potential has a "veto power" over the act that has already been unconsciously willed, and that this veto power is not predetermined and is the source of free will. Although (conveniently) there currently is no way to measure the physiological correlates of veto power, Libet posits that we know it must exist because of the self-reports of the subjective experience of subjects. He appeals to the sense that we can change our minds after having just begun to execute an act. Obviously, a hitter in baseball must be able to check a swing much more quickly than the .5 second interval between the readiness potential and the actual swing. Libet concludes that an immaterial mind may still have a veto power over the physically initiated readiness potential.

Libet goes on to say that, if this veto power *itself* were physiologically predetermined, then free will would be an illusion, and our conscious choice would have no causal power of its own, but would be a mere epiphenomenon. He ignores the possibility, of course, that the choice itself could have its own physiological correlates at the point when it occurs, and thus could have the same causal powers that these physiological correlates have. He seems to assume here that, if A causes B, and then B causes C, this means that B didn't "really" cause C, since A "really" did. In short, the assumption is that a *contra-causally free* will is the same thing as a "will" *per se*. Libet therefore ends up endorsing a straightforwardly *dualistic* conclusion.

An alternative explanation of the Libet findings would be that the readiness potential correlates not with the decision to execute the action, but with a consideration of whether to execute that action, or an imagining of the action that is being considered, *in order* to then decide about it. According to brain imaging studies (Jeannerod, 1994, 1997), imagining an act requires sending efferent action commands, but then inhibiting them. So what the readiness potential is actually measuring may be the imagining of the action, which already involves motor cortex and sensory motor area activation. It may not correlate with the *decision* to act or to the activation of the action at all.

Such a position would be consistent with a dynamical systems solution to the mental causation problem, which in my view allows that the person can have the power to control the interaction of micro-components, provided that the person's ontological status can be understood in terms of patterns of physical self-organization which do not contradict ordinary causal laws. I shall now consider what is needed to reconcile the notion of self-causation in the sense of self-organization with the requirement for causal closure of the micro-level components of self-organizing systems.

What Are Dynamical Systems?

A recent approach that purports to accomplish the long-coveted feat of resolving the mental causation problem is the application of dynamical systems theory to the mind–body relation, especially to the most intractable aspect of the problem, the "hard problem" (Chalmers, 1995). Chalmers argues that if we can show that certain physico–chemical antecedents cause the raising of my hand, and that they operate according to the same physical and chemical principles as in non-conscious parts of nature, then giving a complete physical explanation of all such brain events would still leave out of account anything that would explain why there is *consciousness*. By means of the physical antecedents, we can explain only why physical consequents must occur — not why those *physical* consequents would have the property of consciousness. Dynamical systems theory can answer Chalmers' objections to physicalism by showing *why* only certain types of physical systems — complex dynamical ones that include emotional motivations (which of course can sometimes be unconscious) — can have consciousness (Ellis, 1999a, 1999b, 2000a, 2000b; Ellis and Newton, 1998b, 2000a, 2000b).

Our present discussion will be complicated by the fact that there are currently different notions of what a dynamical systems theory consists of. Some dynamical systems approaches (for example, Kelso, 1995) have no quarrel with the reduction of self-organizing systems to conglomerates of one-way, bottom–up causal sequences, so that the pattern of the self-organizing system itself can be viewed as merely an epiphenomenon of the way the micro-level constituents behave, which just happens to work out in such a way as to self-organize the system as a whole. Newton's characterization of the property of self-organization in complex dynamical systems would be consistent with this reductionist viewpoint:

> The self-organizing properties of complex systems . . . incorporate a natural tendency toward order, which arises spontaneously among molecules when in sufficiently complex groups, in a way that can be explained entirely by physical mechanisms and

> involves no mystery. It does, however, allow the emergent order to be conceptually distinguished from the substratum in ways that appeal to some theorists. Mental states are not reducible to the individual states of the substratum, but they are physical states nonetheless, and obey physical laws. (Newton, 2000, p. 91)

But another trend in self-organizational theory (for example, Anderson and Mandell, 1996; Kauffman, 1993; Monod, 1971; Weiss, 1968) would hold that when a dynamical system is complex enough, not only is it multiply realizable with respect to its physical substrata, but it can also play an active role in *seeking out, appropriating, replacing, and reproducing* the substrata that are needed to maintain the organizational pattern of the system. There seems to be controversy as to whether the self-organizing system has real causal power independently of the causal powers of its separate micro-level constituents, and I shall return to this question.

Dynamical systems theorists flesh out the notion of self-organization in terms of open thermodynamic systems, of which biological organisms are examples. Open thermodynamic systems continuously exchange constituent components and energy with their environment, yet maintain homeostatic constancies across these exchanges; these constancies preserve continuity of structural organization into the future. A behavior pattern into which the system has a strong tendency to settle is called an "attractor" or "basin of attraction." The organism learns and remembers new perceptual patterns by creating new basins of attraction structurally related to the learned stimulus pattern (Alexander and Globus, 1996; Freeman, 1987, 1988). In Freeman's work on olfactory learning in rabbits, as summarized by Alexander and Globus, "There are different basins of attraction, in the form of limit cycles, for different odors the rabbit can recognize . . . . Upon presentation of a *novel* odor, the olfactory bulb is pushed into chaotic activity [allowing] formation of a new limit cycle attractor to suit the novel stimulus" (1996, p. 42).

For present purposes, we can think of a dynamical system as an open thermodynamic system that exchanges energy and materials with its environment while maintaining continuities of structure at a level of organization higher than the level of the substratum components that are continually being appropriated and replaced by the system, and is causally robust with regard to the various inputs into the system. For example, Kauffman (1993) argues that self-organization occurs spontaneously given a sufficient diversity and number of entities capable of a sufficiently large number of potential chemical reactions and an autocatalytic structure. A self-organizing system that can catalyze the reactions that maintain its own existence is a "collectively auto-catalytic system." This requires that a system not be too rigidly organized, but that it have different basins of attraction that it can shift into and out of depending on the need for self-maintenance. It can be adaptive for a system to respond to fairly subtle environmental changes with global shifts from one basin to another, and the flexibility of this shifting is thought to be facilitated by a continual proximity to chaos.

Part of the appeal of this notion for purposes of the problem of mental causation is that it may offer resources with which to clarify the relationship between the causal power of the organizational pattern of a system on the one hand, and on the other hand the specific causal powers of specific components of the system (due essentially to the way in which it is multiply realizable). It may be possible that in such a system the overall pattern has causal power over its own components, which it appropriates to subserve the pattern, yet without violating the principle of causal closure at the level of the interactions of the components themselves. Monod (1971) and Kauffman (1993) specifically claim that this concept of a dynamical system with a structural tendency to maintain itself as a whole across a divesity of components and inputs does not contradict the normal causal laws that constrain the discrete interactions of the components of the system, but rather supplements them with a different kind of analysis — a structural analysis — that is also needed for a complete explanation. If so, then such an approach to the phenomenon of self-organization could help to address the problem of mental causation by attributing a top–down causal power to states of consciousness, which could be identified with structural or relational properties of the system; these structural properties would have the power to maintain themselves across multiply realizable replacements of their physical substrata. This claim is made only for very complex dynamical systems – complex enough for the pattern itself to seek out and replace the components needed to maintain the pattern.

We can thus think of a dynamical system as one whose organization creates a strong tendency to maintain itself across various alternative causal mechanisms at the level of the components whose higher-order structural relations instantiate the system. In a complex dynamical system, not only is the structural pattern multiply realizable with respect to alternative sets of substrata, but it also plays an active role in bringing it about that one or another of the combinations of substrata needed to maintain the overall pattern will obtain. Kauffman and Monod define this dynamical relation in terms of both self-maintaining and self-organizing systems. A self-organizing system not only has a strong tendency to maintain its pattern in the way just defined, but also has a strong tendency to come *into* being in the first place. All self-organizing systems are self-maintaining, but the converse does not necessarily hold. For purposes of applying the theory to the mind–body problem (for example, Edelman, 1989, 1992; Freeman, 1987; MacCormac and Stamenov, 1996; Thelen and Smith, 1994) we can assume that the dynamical system (the biological organism) already exists, and need only worry about how the system is *self-maintaining* across multiply realizable substrata, some available subset of which the system actively seeks out, appropriates, replaces, and reproduces. In speaking of "biological organisms," there is no assumption that artificial systems could not manifest such structures, thus in principle that such systems could not have conscious minds. The determinative factor is not whether the system is composed of certain specific elements such as silicon or carbon, but rather the structural dynamics of the system.

Some of the theorists for whom complex dynamical systems can have causal power over their constituents explicitly contrast dynamical systems against simpler connectionist systems by virtue of emphasis on the top–down causal role of the higher-order pattern that not only maintains itself across replacements of its components, but seeks out and organizes usable components. Alexander and Globus are particularly straightforward about the implications of their view of dynamical systems for the causal analysis of events. "[In] connectionist models . . . transformations are carried out by individual neurons . . . by receiving weighted activations from other neurons" (1996, p. 32). By contrast, in agreement with Freeman (1975), Globus (1992) and Nicolis (1986), Alexander and Globus characterize edge-of-chaos dynamical systems in this way: "[When] interconnectivity *within* a particular scale of organization reaches a critical limit, that scale of organization becomes a module in a larger scale of organization . . . [leading to] cascade effects whereby changes at one scale of organization can modify other scales in an exploding chain of reactions up and down the multi-scale structure" (1996, p. 38). The higher level of organization is claimed by such theorists to "constrain" what can possibly occur at the lower level.

How do these different "scalar" levels interrelate causally? A favorite analogy, originally drawn by Harth (1983), is the relationship between laminar and turbulent flow in fluids. "In turbulent mode the macro-scale turbulence is an expression of microscopic variations in the structure of the flow. In the stable flow, this between-scale communication is reversed . . . . 'Laminar flow' refers to fluid moving in an ordered fashion . . . " (Alexander and Globus, 1996, p. 42). In laminar flow, the overall structural pattern "constrains" the causal relations at the molecular level. The way global wave forms in the brain constrain the discrete interactions of their components is supposed to be similar: "Freeman (1988) calls the highly stimulated state *wave mode*, indicating the dominance of the global wave-form . . . . The wave-form now causally constrains the activity of the individual neurons. The wave mode is analogous to the laminar phase in the flow metaphor. The global wave-form constrains the interactions of individual neurons" (Alexander and Globus, 1996, pp. 49–50).

A problematic question is immediately suggested by this example: both turbulent and laminar patterns describe causal interactions between water molecules, with each molecule's behavior explainable in terms of the behavior of others. In principle, each molecule's behavior can be completely explained in this way, without any reference to the overall pattern of flow; it thus seems that, if there is causal closure at the molecular level, the overall pattern can add nothing to the causal explanation of any molecule's behavior that has not already been explained at the molecular level. If so, the causal power of the dynamical system, if not literally identified with the sum of the causal powers of its micro-level constituents, at least reduces to just another epiphenomenon of the sum of the behaviors of the constituents. So if such a model is applied to the mind–body relation — especially to the relation between consciousness and its neurophysiological substrata — it entails just another form of epiphenomenalism, or alternatively a reduction by identification of the causal power of the dynamical system to the causal powers of its micro-level constituents. The mental decision to raise my hand still has no causal power beyond the causal powers of its separate physical components. Thus, if causal closure is not rejected, the decision can have causal power only if reduced to the causal powers of its separate physical substrata, which already exhaust the causal power of the system. But reducing the causal power of the conscious decision to that of its individual physical substrata likewise does not help us understand in what real sense a unified agent or conscious state can have the power to rearrange its own constituents. It is also important to note here that rejecting causal closure would also be an unhelpful move. To reject causal closure would not just mean rejecting the notion that every physical event has a physical cause; it would also mean denying that even those physical events that *do* have causes have *physical* causes.

This is not the result that the top–down variety of dynamical systems advocates want, nor can it help anti-epiphenomenalists avoid the illusory-choice model of mental causation. What is needed is that the structural pattern of the system should make a causal difference to the maintenance of the system, independently of the existence of any *particular* components at the substratum level, as long as suitable components can be found when sought by the system. In philosophy of mind especially, the top–down type of dynamical systems theorist wants to be able to explain phenomena that presumably are not exhaustively explained at the substratum level; for example, such theorists want to deny that mental phenomena are exhaustively explained by connectionist systems in which transformations are carried out by individual neurons. What would dynamical systems theory have to do to make possible some sort of causal power for the organizational structure of a system that cannot be exhausted by its substrata, yet also does not violate physical causal closure?

How Can Top–Down Systems Avoid Violating Causal Closure?

Suppose we assume, as commonly done in theory of causation (for example, Ellis, 1986, 1991, 1995, 2000c; Mackie, 1974), that no causal antecedent can be necessary or sufficient for its consequent except *given certain background conditions* which are presupposed by the causal relation. For example, flipping a switch causes a light to come on only if certain background conditions are in place — good bulb, wiring, etc. We can then formulate the problem of mental causation for dynamical systems in the following way: How can the system have an overall relational pattern (R) such that R has any effect on the causal relations of the physical substrata, P_1, P_2, etc., where P_1 is sufficient to cause P_2 under the given background conditions?

A promising strategy is to consider the possibility that R is structured so as to have a strong tendency for various components of the system to rearrange themselves such that, if the existing background conditions are not conducive to P_1 causing P_2 under those conditions, then the system becomes rearranged so that P_3 will occur, and will cause P_4, which in turn can subserve R just as well as P_2 could, because of R's multiply realizable nature. If the causal power of the system as a whole is explained in terms of its being structured so as to have a tendency to rearrange the given background conditions for any discrete causal relation within the system, then it becomes comprehensible how the system as a whole, by virtue of its structure, can actively replace one discrete causal process with another, according to what is needed to maintain the structure of the system, without violating the causal closure of the discrete causal relations between the components. To say that the system rearranges the background conditions for a specific causal relation is to deny neither the causal sufficiency of that relation itself (under appropriate background conditions) *nor* the *previous* causal determination of the genesis of the self-maintaining system itself. Nowhere would physical causal closure be violated.

Someone might worry that the logical possibility of such a system does not show that there *are* any systems of this kind, let alone that consciousness can be conceived of as an aspect of such systems in relation to their physical substrata. But whether there are any such systems is an empirical question. Kauffman, Monod and others cite numerous examples of biological organisms that seem to behave according to a principle of self-organization in the sense that the pattern shows a strong tendency to appropriate needed substrata for the maintenance of the pattern. When a given mechanism for maintaining a 98-degree body temperature fails, the organism finds some alternative way to achieve it. When the victim of a mild stroke has lost the neural substrata for certain conscious functions, new cells are appropriated to subserve the relational behavior of those that were destroyed. Now these examples do not prove that the relevant process in each example is exerting real causal power to appropriate and replace constituents, rather than being caused by the discrete actions of the constituents, as top–down dynamical systems theory would have it. All they show is that there are patterns that do act systematically in such a way that, in the final analysis, the continual replacement of parts needed for the pattern's continuity does tend to be achieved in a very robust way, and with a great degree of flexibility with regard to initial conditions. And a dynamical systems model that coheres with all relevant causal facts would be a plausible way of accounting for this robust continuity of the pattern across such a broad range of substratum replacements.

A more positive reason for believing that some systems do fit the dynamical systems causal analysis is that a dynamical systems hypothesis would be consistent with the observed fact of mental causation in a way that no other kind of causal analysis seems coherently able to accommodate. But that depends on whether one grants the arguments cited at the outset that the illusory-choice model is implausible. If dynamical systems theory can account for the phenomenon of mental causation in a coherent way, whereas competing accounts entail an illusory-choice model, then dynamical systems theory would have a decided advantage over the competing theories.

A more pressing problem is whether the top–down causal role assigned to a process over its own substratum elements can even *be* a coherent causal account in the first place. Making such a possibility into a coherent theory requires some conceptual tools if such a scenario is to be clearly distinguished from one in which the discrete behavior of the constituents simply causes the pattern of the process, as in most connectionist systems. To establish that such a theory could cohere with a reasonable causal scenario is the task to which we must now turn.

To make the idea of self-maintaining systems as determining background conditions of causal relations a little more explicit, we can say that a self-maintaining system is a system whose organizational pattern, R, is such that, given any physical substrata for R (P_1 or P_2 or P_3 or . . . etc.),

> (1) if P_1 causes P_2, P_1 *all by itself* is neither necessary nor sufficient for P_2 given R, although there are certain possible background conditions, BC, such that, given BC, P_1 *is* or *would be* necessary and sufficient for P_2, depending on whether BC obtains or not;

and

> (2) the causal relations that actually occur in the system have a strong tendency to combine in such a way as to guarantee that pattern R will continue to obtain in the future.

Turning our attention to the crucial point here — condition (1) — we see that in any self-maintaining system R, it is not the case that, if A causes B, then A iff B simpliciter. What *is* true is that

(I) Given R, $(A_1 \vee A_2 \vee A_3 \vee \ldots)$ iff $(B_1 \vee B_2 \vee B_3 \vee \ldots)$

In ordinary language, R has the ability to manipulate the background conditions needed for any micro-level causal relation such as A_1 iff B_1, A_2 iff B_2, etc., so that at least one of these causal sequences is likely to occur. And at the same time,

(II) Given R, $(A_1 \vee A_2 \vee A_3 \vee \ldots)$ will obtain in enough instances to ensure the continuation of organizational pattern R in the future.

It is true that there is some possible set of background conditions, BC_1, such that

Given BC_1, A_1 iff B_1

and there is some possible BC_2 such that

Given BC_2, A_2 iff B_2.

And

Given BC_3, A_3 iff B_3.

But in a self-maintaining system,

$$R \rightarrow [(BC_1 \vee BC_2 \vee BC_3) \; \& \; (BC_1 \text{ iff } A_1) \; \& \; (BC_2 \text{ iff } A_2) \; \& \text{ etc.}]$$

in enough instances to ensure the continuation of R. That is, R ensures (in enough instances to keep the system going) that if A_1 is available to cause B_1, the needed background conditions will be arranged to facilitate this sequence (that is, BC_1 will obtain); but if A_1 is not available, whereas A_2 is, then the background conditions (BC_2) will be arranged so as to facilitate the sequence $A_2 \rightarrow B_2$, which in turn can subserve the relation R just as well as the sequence $A_1 \rightarrow B_1$ could have done.

Thus, in a self-maintaining system, R ensures that both (I) and (II) above obtain in enough instances to ensure the continuation of R. In most physical situations, which are not self-maintaining, this is not the case.

In every instance where A causes B, there are some given background conditions under which A iff B, but these background conditions by themselves do not constitute a self-maintaining system. The fact that, given certain background conditions, A iff B, does not make this set of background conditions equivalent with a self-maintaining system, although these background conditions may occur *within* a self-maintaining system. In sum, a self-maintaining system is organized in such a way that if the background conditions do not obtain under which A_1 iff B_1, then there is a very strong tendency that the background conditions under which A_2 iff B_2 *will* obtain, or the background conditions under which A_3 iff B_3 will obtain, or etc., as a result of the organizational structure of the system.

This notion of a "strong tendency to combine in such a way as to guarantee that pattern R will continue to obtain" needs to be fleshed out a little. There is an important question as to *how* a self-organizing system can exercise such a tendency. But there are really two different questions packed in here. One is whether there are any such systems, and the other is, if there should be such systems, how it is possible for them to act in this way. The first question is an empirical one, and can best be answered by means of concrete examples, as already mentioned. When the cells of an embryo are transplanted from one brain area to another, sometimes even in a completely different species, at certain stages of development, they take on the functional properties of the new brain area in spite of the alien origin and function of the original cells (Kandel and Schwartz, 1981). In effect, the functional organization of the system appropriates the alien cells for its purposes. In the same way, when stroke victims attempt to use a paralyzed limb during concerted amounts of time over a period of several weeks, different brain cells and synapses are appropriated to serve the function of the old destroyed cells and synapses. If this effort is not exerted in this way, the new cells are not appropriated to serve the lost functions. These examples illustrate that there are instances where functional properties of a larger system appropriate micro-components as needed — within certain limits, of course — rather than merely resulting from the interaction of the micro-components.

The second part of the question is *how* self-organizing systems accomplish this purpose. The main principle seems to be that the pattern of the system, multiply realizable by different possible components of this pattern, causes flexibility in the arrangement of the background conditions needed for any given causal micro-sequence to obtain. In effect, the design of the system includes conditions that allow a robust array of antecedents for massive overcausation of a certain outcome. This can be seen, for example, in the shunt mechanisms built into the Krebs energy cycle. Because the system is patterned to contain these mechanisms for overcausation, there is no one element of the system that is either necessary or sufficient for the final outcome — the conversion of energy into ATP for storage. Instead, the initial structure of possible shunt mechanisms is necessary and sufficient to ensure the outcome, provided that minimal alternative chemical substances that are usually readily available can be obtained. Natural selection will favor the survival of systems that are structured in this way, and a biochemical structure including auto-catalysis further increases the probability that this kind of mechanism will obtain. If none of the alternative substances is indeed available, of course, then the outcome is not produced, the system fails (dies), and ceases to be a self-organizing one.

It might be argued that even though R is not identical with any of the Ps that subserve it, there is a relation, R*, at the level of P (in some sense), such that R = R*. For example, if the Ps are at the level of nerophysiology, then R might be identical with a relation that is neurophysiological in nature. Or if the Ps are at the level of elementary physics, then R might be identical with a relation that is the referent of some complex predicate of elementary physics. In principle, any brain process should be describable at any of these levels, although in practice the ability to do so would depend on whether the needed empirical observations can be carried out at the needed levels, and it cannot be assumed automatically that they can be (see Ellis, 1999a). But the important point for now is that the describability of R in terms of R* would not erase the distinction between the causal powers of R (and of R*) and the causal powers of the Ps that serve as the substrata for the R (and for the R*).

Is R, then, a "physical" relation? What we have seen is that R *could* be physical without thereby having its causal power reducible to the sum of the causal powers of its micro-level components. However, I should not claim to have *shown* in this essay that R is definitely physical, which would be too ambitious an undertaking. What I want to claim to have done is to make possible the juxtaposition of an irreducible R with Kim's causal closure. It may be that, if R is not physical, then causal closure would still be violated. But we must be somewhat cautious on this point, because there are different definitions of "physical."

In the most frequently used sense (the one used by Kim, for example), "physical" simply means whatever can be studied by means of the physical sciences. But this is a problematic sense, because there is the possibility that something can be for one reason or another inaccessible to the methods of those particular sciences, yet still be a process that is causally attributable to what is accessible to the methods of those sciences — which would make them "physical" in a different sense, a broader and "derivative" sense. In that case, one could say that something is physical if either (a) it can be studied by the methods of the physical sciences, or (b) it causally interrelates with things that can be studied with the methods of the physical sciences. The danger of not including phenomena of type (b) in the definition of "physical," as I discussed more extensively elsewhere (Ellis, 1999a), is that the more limited definition seems to conflate "physical" with "empirically observable." There are many problems with this definition, not the least of which led to the demise of the logical empiricism of the mid-twentieth century. There are simply too many scientific entities that can be inferred from what is observable, yet are not themselves empirically observable. But this does not stop them from fitting into the causal structure of a physical theory.

In principle, to avoid confusing epistemological with ontological claims, there should be the logical possibility of hypothesized physical entities that cannot be directly empirically observed, although their existence may be inferrable from what is empirically observable. In this broader sense, then, and with the qualification that some physical events may not be directly observable, we can say that the R that subserves conscious choices could quite well be a physical relation, even though its causal powers are not reducible to those of the components of R that *are* observed with the methods of the physical sciences. But, again, it would be beyond the scope of this paper to purport to prove that R definitely is physical. What I have shown is that resolving the problem of mental causation does not require that it be non-physical. It seems clear that the R we have been referring to is conceptualized, from the standpoint of dynamical systems theory, as a physical R, and that there is no reason why it cannot be physical while at the same time its causal power remains irreducible to the sum of the powers of its components.

Objections and Responses

Against this argument for the compatibility of ordinary causal laws with the causal power of a self-maintaining system over its own constituents, the following objection might be raised: to make this case, it might be argued, such a top–down type of dynamical systems theory must show that it is the "overall relational pattern of the system (R)" that actually *drives* or *causes* the rearrangement; but the objector might hold that it is the causal relations and interactions at the micro-level that drives the rearrangement, and that the overall relational pattern of the system is just a higher-level, coarse-grained description of what is going on at the micro-level. What is really doing the causal work, in the case of mental causation, is "down there in the neurons."

The response is that, first of all, the relata for a relation (in this case, the Ps) could have occurred without being in that relation. So the R is distinguishable from the Ps in that it could have been the R it is, and thus could have created the background conditions needed for subsequent effects, without the particular Ps that, under the given circumstances, happen to subserve it — as long as other suitable Ps could have been found. This means that whether or not R obtains makes a difference in determining subsequent causal relations, while R's making this difference is not dependent on the action of the *specific* Ps that, under the given circumstances, happen to subserve it. Other Ps not only could have, but in fact *would very likely* have subserved the same R if those Ps had not, and it is the R that is necessary, not those specific Ps.

Secondly, it is true that the R ultimately is determined by some Ps (at the micro level), but the R is determined by *previous* Ps, not by the Ps that serve as the substratum for the current R. Therefore, the R isn't just a courser-grained description of what is happening in *its* micro-level substrata; the R has the causal power to appropriate different micro-level substrata as needed at each subsequent moment, by rearranging the background conditions to facilitate available causal sequences capable of subserving the ongoing pattern at that moment. Thus the R's causal power is not literally identical with the causal power of its micro-level substrata, since it is counterfactually true that the R could have *and very likely would have* had that causal power without those particular Ps, although of course the R's causal power is accountable for by the causal mechanisms that caused that R to be organized in the way it is in the first place; and *those* causal mechanisms are ultimately (at some point) traceable to *previous* micro-level events.

In case this sounds like a trivial point (that all it means is that at each instant in time the operative R at that time is caused by the Ps in the immediately previous instant), notice that it means much more than that: it means that not all of the causal work is being done "down there in the neurons." It is also being done by the *way in which the neurons are organized*, which is a higher-level relational property of the system irrespective of which micro-level constituents are serving as the relata for the relation.

I have claimed that the current R, caused by previous Ps, is doing indirect causal work *now*, in the sense that it is rearranging background conditions needed for the Ps to be causally sufficient for each other. I have taken this to imply that the current R is not identical with or caused by the specific current Ps, having been caused by previous Ps. But there still might be a problem with this answer: Why couldn't one still hold that the previous Ps caused the current Ps, and that it is the current Ps, not the R, that are doing the causal work?

Part of the question here hinges on what is meant by "causal work." When R rearranges the background conditions under which one P will or will not be necessary and sufficient for another, is it doing "causal work?" Well, yes and no. It isn't creating any situation in which there is any P that doesn't have another P that is necessary and sufficient for it, under the given background conditions (and of course *nothing* does *any* causal work except under given background conditions!). But what it is doing makes a *counterfactual* difference, in that, if the R hadn't arranged the background conditions in the right way, then the first P *would not have* been necessary and sufficient for the next. The important point for our purposes is that the R *makes a difference* as to whether a given P is able to do the causal work assigned to it or not.

Dynamical systems theory does not need to claim that the R itself isn't in turn caused by previous Ps.

The other part of the problem, of course, is the question of the ontological status of R. We have seen that R cannot be literally identical with its *currently subserving* Ps, since a relation is not the same as the relata that are in that relation, and there are counterfactual statements that are true of the R but not of the specific relata that subserve it; most importantly, the R is necessary for subsequent effects for which those Ps are *not* necessary, since if those Ps had not occurred in relation R, others would have, because of the way R was structured. Thus there is a counterfactual statement that is true for R but not for the Ps that subserve the R, namely that the R is necessary for certain subsequent occurrences for which the Ps would *not* have been necessary, if other suitable Ps had been the ones to subserve R.

If the R were literally identical with the Ps that are in relation R, then the R and its subserving Ps would both do all the same "causal work"; but we have seen that the R is not literally identical with its Ps. Moreover, the R cannot be literally identical with the *previous* Ps that caused it, because things that don't exist at the same time cannot be literally identical; and also, if X causes Y, then X and Y cannot be literally identical with each other. So R is literally identical neither with its *currently subserving* Ps nor with any *previous* Ps, although it might be identical with some *relation* between the two — but the causal work done by that relation itself is not reducible to the causal work being done by the Ps, for three reasons: (1) a relation, by definition, is not the same thing as the relata that are in that relation; (2) the R doesn't exist *at the same time* as the Ps that caused it; and (3) as we have seen, there are counterfactuals that are true of R that aren't true of the Ps and vice versa. For example, R could have existed without the currently subserving Ps, if other appropriate Ps could have been found.

The notion that R involves relations between the previous Ps and the current, subserving Ps seems like a very promising line for dynamical systems theory to take; it helps in answering the following concern: consider the presupposition raised by the above question "When R rearranges the background conditions . . . is it doing causal work"? The presupposition seems to be that R does rearrange the background conditions. But what grounds can we have for saying that R itself "*does*" anything? The question is: Is R itself doing the rearranging? Or, on the contrary, are the Ps that subserve the R actually doing the rearranging? If we say that R is really doing the rearranging, what are the grounds for saying that it is?

Phrased in this way, the question becomes what we mean by "doing" in "doing something." What the R "does" is to be a necessary and sufficient condition for an outcome for which no set of actually occurring Ps is necessary; nor could those Ps have been sufficient for that outcome without that R. That is, if the Ps were arranged in some different R, then there would be a different outcome, and the R in question not only *could* but *would* have led to that outcome without those Ps if they had not been available (because R is structured so that other Ps would have ended up serving as its substrata in that case).

Someone might respond here that the fact that the Ps are arranged in this R is itself a P — a higher-order P. But that would just be another way of stating the top–down dynamical systems theorist's point — that we cannot exhaustively explain the higher-order relation in a way that completely reduces to the explanations of the lower-order Ps that subserve it, although of course we *can* explain the higher-order relation in terms of some *earlier* lower-order relations (but again, this is not the same as literally reducing it to them, since it doesn't exist at the same time as them, and since there are counterfactuals that are true for it that aren't true for them). So Francescotti (1998) is right to insist that multiple realizability is not an adequate argument against *physicalism*, because the R can be a physical R even if it is multiply realizable; but in the context of a situation where the multiply-realizable process is self-maintaining, the multiple realizability of the process can be an argument against a (physical) R's reducibility to the Ps that are *in* relation R. R can be a physical R without thereby being reducible to the micro-level Ps that subserve it.

This point illustrates the importance of the distinction Weinberg (1995) has made between what he calls "petty reductionism" and "grand reductionism" in science. Petty reductionism is defined as the view that "things behave the way they do because of the properties of their constituents." By contrast, "Grand reductionism . . . [is] the view that all of nature is the way it is because of simple universal laws [which would] reduce the world of physical phenomena to a finite set of fundamental equations" (p. 39). According to these definitions, Weinberg says that "petty reductionism in physics has probably run its course" (p. 39). Dynamical systems theory, in order to make sense of the problem of mental causation, does not need to reject Weinberg's grand reductionism — only what he calls petty reductionism.

A final objection to this way of conceptualizing a self-maintaining system is based on metaphysical considerations: just because a system continues to display the same structural patterns across time, this does not permit us to conclude that numerically *the same* system persists across time. So, metaphysically, the components may be what are "real," and the self-maintaining system may be merely an epiphenomenon of component actions.

There are two answers here. First, to declare that the self-maintaining system is "merely an epiphenomenon" is not to say that the self-maintaining system as just defined does not exist, nor that it does not have the tendencies described, which give it the power to maintain itself across replacements of its substrata, thus allowing a meaningful distinction between the system and its components. What the objection really entails, if its premise is true, is that self-maintaining systems have their own causal antecedents, just like everything else in nature. But to say that they are caused to be the way they are does not negate the possibility that they in turn can play a role in bringing about subsequent effects.

Secondly, it is not necessary for self-maintaining systems to maintain numerical identity across time in order for them to play the role that top–down dynamical systems theorists want to assign them. That would be necessary only if we were interested in the problem of the continuity over time of personal identity. If "I" am a pattern of organization, then does the selfsame "I" continue to exist when a similar pattern obtains in the future, or does a different "I" exactly similar to myself exist? This is a stubborn metaphysical perplexity that accompanies any conceptualization of personal identity, and is not unique to dynamical systems theory. For present purposes, it can be postponed to another time.

The important point for now is that it is the *past* background conditions that determine the *present* state of the system, and that systems (for example, the Krebs cycle) have *evolved* to select environmental elements or states that will enable certain possible alternative requirements to be met in the future (because the ones that didn't adjust the environment appropriately died out). My claim is not that the organism at a particular time forms its background conditions at that same time in order that it can be in the state it is actually in at that time. That would be an incoherent claim. Self-maintaining organisms are forward-looking (which is why some of them evolved representational systems), so they are able to act now so that their future needs will be satisfied. And these future needs can be needs of the organism as a whole, not just of the individual components in their current configuration. But it can still be the case that the components are constructed, by this evolutionary selection process, to interact with the environment to produce conditions that will turn out to be (probably) advantageous. Dynamical systems offer an understanding of a general mechanism that could produce this result. Does this process involve a magical ability of an organism, existing at time T, to construct itself during that same time T? Not at all: the process takes place over an extended period of time.

Conclusion

To say that one thing causes another is to say that if the antecedent in question had occurred under such-and-such background conditions, then a certain consequent would have followed, whereas if the antecedent had failed to occur under similar background conditions, then the consequent would have failed to follow (Ellis, 2000c; Mackie, 1974). The asymmetry of the causal relation can be accounted for on this interpretation because the antecedents would have been sufficient to guarantee the outcome, given certain background conditions, and it is also true that if those sufficient preconditions had failed to occur under similar background conditions, then the consequent would have failed to occur under those background conditions. But, since the background conditions under which the antecedent could have failed to occur are somewhat different from those under which the antecedent does occur, the background conditions for the statement "given certain background conditions, A is sufficient for B" are different from the background conditions for the statement "given certain background conditions, if A had not occurred under those conditions, then B would not have occurred." This difference in the background conditions needed for the two relations between A and B guarantees the causal asymmetry between A and B.

Given this account of traditional causal laws, suppose now that the choice to execute an action is necessary and sufficient under the relevant background conditions for the execution of that action. So, if I had not decided to raise my hand, it would not have gone up, and anytime I should decide to raise it under essentially similar background conditions, it would have gone up. In the top–down type of dynamical systems approach to consciousness, the agent who makes the decision is a relation, requiring suitable physical substrata, and the conscious choice to raise one's hand is another such relation, which the above discussion designated as R. This R is multiply realizable by a number of alternative sequences of causal micro-mechanisms, with each alternative sequence requiring its own set of background conditions. R is necessary and sufficient for the hand's going up, whereas no particular set of Ps is necessary and sufficient for the hand's going up except given certain background conditions which are substantially controlled by R. We are assuming that, in a suitably complex dynamical system, R could have obtained without that particular set of Ps, by rearranging the background conditions in such a way that a different set of Ps could have subserved R. Thus, by the above definition of causation, R causes the hand to go up, yet we can also say that, given the background conditions that did in fact obtain, the Ps that subserved R also caused the hand to go up. But if the Ps that caused the hand to go up had been unavailable to do so, R could have rearranged the background conditions so that, under these altered background conditions, an alternative set of Ps could have caused the hand to go up.

Not only can R meaningfully be designated as a cause of the hand's going up, but in fact R is a cause of the hand's going up under a broader set of possible background conditions than any given sequence of Ps. Moreover, R is both multiply realizable by alternative sets of Ps, and has some control over the realization of alternative sets of background conditions under which a given sequence of Ps will or will not be used as the particular sequence leading to the raising of the hand. Thus R has more control over the raising of the hand than any particular set of Ps subserving R can have, since it can be said that R is necessary and sufficient for the hand's going up under many background conditions where that particular sequence of Ps would not have sufficed to make it go up (because the needed background conditions may not have been in place).

The causal power of R therefore is not illusory, yet neither does it contradict the causal closure of micro-level sequences, because the micro-level sequence that actually does occur is necessary and sufficient to bring about the resulting action *under the given background conditions*. But these background conditions have already been influenced by a previous enactment of R, whose tendency is to manipulate background conditions in ways needed so that available micro-level substrata can produce the outcomes that will propagate R into the future.

Nor does this causal power of R contradict the thesis that R itself may have been caused to be the way it is by some past set of micro-level events (for example, in ontogeny and phylogeny). The possibility that I may have been caused by a combination of environmental and hereditary factors to be the kind of person who would choose to study philosophy does not contradict the fact that I did choose to study it. If someone writes a will, we may ask for evidence that the person chose to write it to include certain specific contents, as opposed to being coerced by forces contrary to his or her choice, and this question remains meaningful regardless of whether the personality of the person writing the will has been predetermined.

In sum, we do make choices, and choices do cause actions, even though the actions can also be said to have been caused by substratum-level events, if certain background conditions for those causal relations have obtained. However, when a mental choice occurs, the choice is an aspect of a motivated self-organizational process that is a relation subserved by the relevant substratum events, and this self-organizing process itself is structured so as to have a strong tendency to go out and seek the needed substrata to keep its pattern of organization going. The relational process, then, has the power (within limits, of course) to replace its own components, and this power is essentially the basis for the causal power of mental choices over our bodily movements.

References

Alexander, D., and Globus, G. (1996). Edge of chaos dynamics in recursively organized neural systems. In E. MacCormac and M. Stamenov (Eds.), *Fractals of brain, fractals of mind* (pp. 31–74). Amsterdam: John Benjamins.

Anderson, C. (2000). From molecules to mindfulness: How vertically convergent fractal time fluctuations unify cognition and emotion. *Consciousness & Emotion*, *1*, 193–226.

Anderson, C., and Mandell, A. (1996). Fractal time and the foundations of consciousness: Vertical convergence of 1/f phenomena from ion channels to behavioral states. In E. MacCormac and M. Stamenov (Eds.), *Fractals of brain, fractals of mind* (pp. 75–126). Amsterdam: John Benjamins.
Chalmers, D. (1995). Facing up to the hard problem of consciousness. *Journal of Consciousness Studies*, *2*, 200–219.
Damasio, A. (1994). *Descartes' error*. New York: Putnam.
Damasio, A. (1999). *The feeling of what happens*. New York: Harcourt Brace.
Davidson, D. (1970). Mental events. In L. Foster and J.W. Swanson (Eds.), *Experience and theory* (pp. 79–102). Amherst: University of Massachusetts Press.
Edelman, G. (1989). *The remembered present*. New York: Basic Books.
Edelman, G. (1992). *Bright air, brilliant fire: On the matter of mind*. New York: Basic Books.
Ellis, R.D. (1986). *An ontology of consciousness*. Dordrecht: Kluwer/Martinus Nijhoff.
Ellis, R.D. (1991). A critique of concepts of non-sufficient causation. *Philosophical Inquiry*, *13*, 22–42.
Ellis, R.D. (1995). *Questioning consciousness: The interplay of imagery, cognition and emotion in the human brain*. Amsterdam: John Benjamins.
Ellis, R.D. (1999a). Why isn't consciousness empirically observable? Emotion, self-organization, and nonreductive physicalism. *Journal of Mind and Behavior*, *20*, 391–402.
Ellis, R.D. (1999b). Dynamical systems as an approach to consciousness: Emotion, self-organization, and the mind–body problem. *New Ideas in Psychology*, *17*, 237–250.
Ellis, R.D. (2000a). Consciousness, self-organization, and the process–substratum relation: Rethinking nonreductive physicalism. *Philosophical Psychology*, *13*, 173–190.
Ellis, R.D. (2000b). Efferent brain processes and the enactive approach to consciousness. *Journal of Consciousness Studies*, *7*, 40–50.
Ellis, R.D. (2000c) Three elements of causation: Biconditionality, asymmetry, and experimental manipulability. *Philosophia*, *29*, 1–21.
Ellis, R.D. and Newton, N. (1998a). *Consciousness and the brain: An annotated bibliography*. http://earthlink.net/~dravita. (Internet site.)
Ellis, R.D. and Newton, N. (1998b). Three paradoxes of phenomenal consciousness: Bridging the explanatory gap. *Journal of Consciousness Studies*, *5*, 419–442.
Ellis, R.D., and Newton, N. (Eds.) (2000a). *The caldron of consciousness: Affect, motivation, and self-organization*. Amsterdam: John Benjamins.
Ellis, R.D., and Newton, N. (2000b). The interdependence of consciousness and emotion. *Consciousness & Emotion*, *1*, 1–11.
Francescotti, R. (1998). What multiple realizability does not show. *Journal of Mind and Behavior*, *18*, 13–28.
Freeman, W. (1975). *Mass action in the nervous system*. New York: Academic Press.
Freeman, W. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological Cybernetics*, *56*, 139–150.
Freeman, W. (1988). Why neural networks don't yet fly: Inquiry into the neurodynamics of biological intelligence. *Proceedings of the IEEE International Conference on Neural Networks* Volume 2 (pp. 1–7). San Diego: IEEE (Institute of Electrical and Electronics Engineers).
Globus, G. (1992). Towards a noncomputational cognitive neuroscience. *Journal of Cognitive Neuroscience*, *4*, 299–310.
Haines, D.E., Dietrichs, E., Mihailoff, G.A., and McDonald, E.F. (1997) Cerebellar–hypothalamic axis: Basic circuits and clinical observations. In J. Schmahmann (Ed.), *The cerebellum and cognition* (pp. 84–110). New York: Academic Press.
Harth, E. (1983). Order and chaos in neural systems: An approach to the dynamics of higher brain functions. *IEEE Transactions on Systems, Man and Cybernetics*, *13*, 782–789.
Jackendoff, R. (1996). How language helps us think. *Pragmatics & Cognition*, *4*, 1–34.
Jackson, F. (1986). What Mary didn't know. *Journal of Philosophy*, *83*, 291–295.
Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*, 187–244.
Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Blackwell.

Kandel, E., and Schwartz, J. (1981). *Principles of neural science.* New York: Elsevier–North Holland.

Kauffman, S. (1993). *The origins of order.* Oxford: Oxford University Press.

Kelso, J.A. (1995) *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, Massachusetts: MIT/Bradford.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research, 52,* 1–26.

Kim, J. (1993). The nonreductivist's troubles with mental causation. In J. Heil and A. Mele (Eds.), *Mental causation* (pp. 189–210). Oxford: Oxford University Press.

LeDoux, J. (1996). *The emotional brain.* New York: Simon & Schuster.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies,* 6, 47–58.

MacCormac, E., and Stamenov, M. (Eds.). (1996). *Fractals of brain, fractals of mind.* Amsterdam: John Benjamins.

Mackie, J.L. (1974). *The cement of the universe.* Oxford: Oxford University Press.

Monod, J. (1971). *Chance and necessity.* New York: Random House.

Nagel, T. (1974), Physicalism. *Philosophical Review, 74,* 339–56.

Newton, N. (1996). *Foundations of understanding.* Amsterdam: John Benjamins.

Newton, N. (2000) Conscious emotion in a dynamic system. In R. Ellis and N. Newton (Eds.), *The caldron of consciousness: Motivation, affect and self-organization* (pp. 91–105). Amsterdam: John Benjamins.

Nicolis, J.S. (1986). *Dynamics of hierarchical systems.* New York: Springer Verlag.

Panksepp, J. (1998). *Affective neuroscience.* New York: Oxford University Press.

Schmahmann, J. (Ed.). (1997). *The cerebellum and cognition.* New York: Academic Press.

Searle, J. (1984). *Minds, brains, and science.* Cambridge, Massachusetts: Harvard University Press.

Smart, J.J.C. (1963). Materialism. *Journal of Philosophy,* 60, 651–662.

Smart, J.J.C. (1970). Sensations and brain processes. In C.V. Borst (Ed.), *The mind–brain identity theory* (pp. 141–56). London: Macmillan.

Spence, S.A., and Frith, C.D. (1999). Towards a functional anatomy of volition. *Journal of Consciousness Studies,* 6, 11–29.

Thelen, E., and Smith, L. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, Massachusetts: MIT Bradford.

Watt, D. (2000). The centrencephalon and thalamocortical integration: Neglected contributions of periaqueductal gray. *Consciousness & Emotion, 1,* 91–114.

Weinberg, S. (1995). Reductionism redux. *New York Review,* Oct. 5, pp. 39–42.

Weiss, P. A. (1968). The living system: Determinism stratified. In A. Koestler and J.R. Smythies (Eds.), *Beyond reductionism — New perspectives in the life sciences* (pp. 3–55) [The Alpbach Symposium]. London: Hutchinson.